DITCHEVA, MARIA, M.A. Trait-like Depression Risk Factors and Cortisol Reactivity to Lab-Induced Stress: A Meta-Analysis Examining Stressor Severity. (2019)
Directed by Dr. Suzanne Vrshek-Schallhorn. 93 pp.

Dysregulated cortisol reactivity has been associated with trait-like depression risk factors; however, findings regarding the direction of the association have been inconsistent, with evidence of both blunted and heightened reactivity. The Cortisol Reactivity Threshold Model aims to reconcile these divergent findings by positing that individuals vary in their cortisol sensitivity to stressors in a systematic fashion, generating peak cortisol reactivity at different levels of stress severity, with at-risk individuals' reactivity peaking at lower levels of threat, and lower risk individuals' reactivity peaking at higher levels of threat. This suggests that stressor severity moderates the risk-reactivity relationship in a curvilinear fashion. In this meta-analysis, I examined the relationship between trait-like depression risk factors (extraversion, negative cognitive style, neuroticism, perfectionism, and rumination) and cortisol reactivity to lab-based stressors in 40 independent experimental samples. Specifically, I tested the Cortisol Reactivity Threshold Model using meta-regression to examine a hypothesis that the risk-reactivity relationship varies as a curvilinear function of stressor severity; I also examined a series of moderators. No significant overall effect size emerged between depression risk and cortisol reactivity aggregating across severities (g=0.039; p=0.609). A curvilinear effect of stressor severity was not a significant predictor of the risk-reactivity association across all risk factors aggregated (β=-0.251, p=0.163) and the effect for trait rumination approached significance (β=-0.593; p=0.062); however, several potentially model-consistent moderator findings emerged. Proportion female (β=-0.598; p=0.038) and

stressor anticipation (β=-0.069; p=0.003), both predicted a negative risk-reactivity association while habituation was associated with a positive risk-reactivity relationship (g=0.541; p=0.007). My confidence in the null stressor severity curvilinear effect findings is tempered by the few studies included with robust stress levels and by the heterogeneity of severity among studies using moderate stress levels. Future research would benefit from examining the effect of more robust lab-based stressors and from standardization of language reporting stressor severity.

TRAIT-LIKE DEPRESSION RISK FACTORS AND CORTISOL REACTIVITY TO LAB-INDUCED STRESS: A META-ANALYSIS EXAMINING STRESSOR SEVERITY

by

Maria Ditcheva

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Arts

Greensboro
2019

Approved by

Dr. Suzanne Vrshek-Schallhorn
Committee Chair

APPROVAL PAGE

This thesis written by MARIA DITCHEVA has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____
Dr. Suzanne Vrshek-Schallhorn

Committee Members _____
Dr. Paul Silvia

_____
Dr. Blair Wisco

_____
Date of Acceptance by Committee

_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

Stressful life experiences have consistently been linked to depression. Additionally, certain trait-like characteristics, such as neuroticism and trait rumination, have been established as risk factors for the disorder. Diathesis-stress theory connects these two lines of research by predicting that increasing levels of pre-existing risk factors confer greater sensitivity to stress, an interaction effect. One way in which this model has been tested is examining how these risk factors predict response to laboratory-controlled stress exposure, primarily focused on reactivity in a stress-responsive hormone, cortisol. However, findings have been inconsistent, with some studies reporting a positive depression risk factor-cortisol reactivity (here "risk-reactivity") relationship and others reporting the opposite, a negative risk-reactivity relationship. The Cortisol Reactivity Threshold Model (Vrshek-Schallhorn, Avery, Ditcheva, & Sapuram, 2018) draws on cortisol's function as a resource-mobilizing hormone in response to stress to explain the divergent findings. Under this model, individuals systematically differ in the threshold of stressor severity that elicits their peak cortisol reactivity. Specifically, the model posits that people at relatively low depression risk exhibit peak cortisol reactivity in response to a robust stressor and a more blunted response to a moderate stressor. However, higher risk individuals exhibit peak cortisol reactivity earlier, under moderate stress, while under a more robust stressor they exhibit a blunted, or potentially insufficient response to the

challenge. The present study is a meta-analysis that aimed to test the Cortisol Reactivity Threshold Model by examining whether stressor severity across studies moderates the risk-reactivity relationship.

Trait-Like Depression Risk Factors

A number of personality and trait-like risk factors have been associated with depression, including extraversion, neuroticism, trait rumination, perfectionism, and negative cognitive style (e.g., Hankin, Abramson, Miller, & Haeffel, 2004; Kotov, Gamez, Schmidt, & Watson, 2010). Relatively stable risk factors such as these are attractive targets for research in part because they may one day be used to identify individuals for preventive interventions (e.g., Barlow, Sauer-Zavala, Carl, Bullis, & Ellard, 2014). Although these constructs have substantial differences, they share in common the tendency to perceive and react to one's environment in a maladaptive manner.

Two personality traits, extraversion and neuroticism, have been implicated in the development of depression. The personality trait extraversion is characterized by a tendency to experience positive affect including high levels of affiliation with other people (Costa & McCrae, 1992); *low* levels of extraversion (often described as high introversion) have been linked with depression risk (Clark, Watson, & Mineka, 1994); though evidence is somewhat inconsistent. For example, a few studies have demonstrated a negative association between depression and extraversion (Fanous, Neale, Aggen, & Kendler, 2007) and depression and positive emotionality, a construct related to extraversion (Naragon-Gainey, Watson, & Markon, 2009). However, a longitudinal study

of female twins found no significant evidence of a relationship between extraversion and depression (Kendler, Neale, Kessler, Heathe, & Eaves, 1993). Moreover, while Malouff, Thorsteinsson, and Schutte (2005) reported a strong link between low extraversion and mood disorders in one meta-analysis, in a subsequent meta-analysis, Kotov et al. (2010) found that dysthymic disorder is strongly associated with low extraversion, however, that major depression was found to be only weakly associated with the trait, indicating inconsistency in findings. This inconsistency regarding extraversion may be due to shared variance with another personality trait, which is associated with depression, neuroticism: A longitudinal study of emerging adults showed that low extraversion significantly predicted depression onset until neuroticism was accounted for (Kendall et al., 2015).

Neuroticism is a personality trait characterized by heightened experiencing of negative emotional states such as anxiety, frustration, low mood, and anger (Costa & McCrae, 1992). This proclivity towards general negative emotionality has been consistently shown to predict the onset as well as the duration of depression (Boyce, Parker, Barnett, Cooney, & Smith, 1991; Clark et al., 1994; Fanous et al., 2007; Kendler et al., 1993). Further support for the association between neuroticism and depression comes from two meta-analyses. One meta-analysis of 33 studies examined the association of the Big Five personality traits with different psychopathologies, and findings indicated that mood disorders were associated with heightened neuroticism (Malouff et al., 2005). A second meta-analysis of 175 studies supported the relationship between neuroticism and depression (Kotov et al., 2010).

A further depression risk factor, trait rumination, is characterized by a pattern of perseverative thinking about one's negative thoughts and feelings, including what caused these and their impact (Nolen-Hoeksema, 1991). Rumination is a complex construct, with several definitions in the literature. For example, Nolen-Hoeksema's (1991) Response Style Theory posits that rumination is characterized by continually focusing on one's depressed cognitions (e.g., "*What is the matter with me that I can't feel better?*"). By contrast, stress-reactive rumination emphasizes the tendency to focus on negative thoughts or experiences activated by stressful or unpleasant life events (Alloy et al., 2000); thus, the focus is on the negative response to the event or the negative event itself (e.g., *"Why am I always the one to have a bad interview experience?"*). Despite this heterogeneity, in general, rumination is hypothesized to contribute to depression by maintaining negative affect and preventing engagement in adaptive strategies, such as problem solving, to alleviate depressed or negative affect (Nolen-Hoeksema, 1991; Nolen-Hoeksema, Wisco, & Lyubomirsky, 2008). Research supports rumination as a risk factor for depression and provides evidence that rumination triggers depression onset (Nolen-Hoeksema, 1991) and predicts duration of the disorder (Robinson & Alloy, 2003). For example, in one study, rumination significantly predicted depression onset at baseline and at a one-year follow up (Nolen-Hoeksema, 2000). In a second longitudinal study, chronically depressed persons exhibited higher rumination than non-clinically depressed persons (Wiersma et al., 2011). Finally, a recent meta-analysis examining emotion-regulation styles in various mental disorders identified rumination as the most potent risk

factor for depression among the various maladaptive emotional responses studied (Aldao, Nolen-Hoeksema, & Schweizer, 2010).

Perfectionism is a personality trait characterized by a rigid tendency to set overly high expectations for oneself coupled with fear of poor or imperfect performance and self-criticism when performance is thought to have fallen short of one's standards (Frost, Marten, Lahart, & Rosenblate, 1990). It is hypothesized that the combination of unfeasible standards in conjunction with critical self-assessment set the stage for perceived failure and contribute to the onset of depression. In support of this, in a longitudinal study, high self-focused perfectionism interacted with achievement pressure to predict depression onset (Hewitt, Flett, & Ediger, 1996).

Finally, negative cognitive style refers to a pattern of maladaptive, negative thinking about events and their interpretations (Hankin et al., 2004). According to Beck's (1979) theory of depression, three-pronged rigid, dysfunctional, negative thoughts about the self (e.g., "*I am a failure"*), the world (e.g., "*Everyone thinks I'm a failure so I have no friends*"), and the future (e.g., "*I will always be a failure and lonely"*) are activated in response to negative experiences and trigger feelings of worthlessness. A further aspect of negative cognitive vulnerability is a negative inferential style, in which individuals make negative inferences about themselves (e.g., "*I am incapable"*), which are global (e.g., *"I failed one test so I will fail all of my classes"*), and stable (e.g., *"I will always fail"*) in nature (Alloy, Abramson, Keyser, Gerstein, & Sylvia, 2008). In Abramson's hopelessness theory of depression, a negative inferential style in response to stressful life events is thought to contribute to the disorder by increasing feelings of hopelessness,

which are hypothesized to be "proximal causes" (p. 358) of the disorder (Abramson, Metalsky, & Alloy, 1989). A prospective study examining non-depressed individuals with high versus low levels of negative cognitive style demonstrated that the former group had increased risk for developing major depression, greater episode recurrence (Alloy et al., 2006; Iacoviello, Alloy, Abramson, Whitehouse, & Hogan, 2006), and a more severe course of the disorder (Iacoviello et al., 2006) than the latter group.

Depression and Diathesis Stress Models

Major depressive disorder (here "depression") refers to the experience of one or more episodes of depressed mood and/or loss of interest or pleasure for two or more weeks accompanied by additional symptoms such as insomnia, loss of appetite, reduced energy, feelings of worthlessness, and suicidal ideation (American Psychiatric Association, 2013). Depression confers the largest public health burden of any mental health condition, and also ranks as a leading health burden among physical conditions due to lost productivity, treatment cost, and lost years of life due to suicide (Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004; Ferrari et al., 2013). Given the high societal costs of depression, understanding its mechanisms of risk represents an important area of investigation.

In addition to the above-described trait-like depression risk factors, which confer increased risk for the disorder, three decades of research have established a robust link between stressful life experiences and various psychopathologies, including depression. Evidence suggests that both chronic and acute stressors are implicated in depressive episode onset (Kendler, Karkowski, & Prescott, 1999), symptom severity (Hammen,

Davila, Brown, Ellicott, & Gitlin, 1992), and maintenance of the disorder (Hammen, 2006). The dominant diathesis-stress model of depression posits that stress activates diatheses, or preexisting risk factors (e.g., trait rumination and neuroticism), to contribute to depression onset (Monroe & Simons, 1991). This model predicts that individuals high in these risk factors are differentially vulnerable to the effects of life stress as compared to those lower in these traits (Kendler, Kuhn, & Prescott, 2004). Thus, examining the relationship between reactivity to an acute stressor and trait depression risk factors may shed light on their role in precipitating depression.

The HPA Axis: Cortisol Reactivity

One approach to examine how these pre-existing depression risk factors operate under stress in controlled conditions is through lab-based stress induction. Much of this research examines response to a lab-based stressor by measuring the production of the stress-responsive hormone, cortisol, the end product of the hypothalamic-pituitary-adrenal (HPA) axis. One mechanism through which stressful life experiences exert their influence on an organism is by activating the HPA axis, (for a review, see Sapolsky, Romero, & Munck, 2000; Tsigos & Chrousos, 2002). In response to acute physiological or psychological stressors, the hypothalamus is activated and releases corticotropin releasing-hormone (CRH). In turn, CRH activates the anterior pituitary gland to secrete adrenocorticotropic-hormone (ACTH), which travels in the blood stream and binds to receptors in the adrenal cortex, which stimulates synthesis and secretion of the end product, cortisol. Cortisol then triggers numerous physiological and cognitive changes to mobilize resources. These changes include elevated heart rate, blood pressure, and blood

glucose levels, as well as heightened attention and memory, especially for emotional content (Smith & Vale, 2006). Several aspects of HPA functioning have received considerable attention in relation to depression, including the cortisol awakening response (CAR, the normative rapid rise in cortisol upon waking; Fries, Dettenborn, & Kirschbaum, 2009), its diurnal rhythm (peaking in the morning and declining throughout the day; Tsigos & Chrousos, 2002), and acute cortisol reactivity to a stressor, the focus of the present study. Cortisol reactivity (also termed cortisol response) refers to the change in cortisol levels from basal levels following an acute stressor, although it is operationalized in several ways with subtle interpretive differences (see Table 1).

Traditionally, heightened reactivity has been considered indicative of a more robust and also maladaptive response to stress (Heim et al., 2000). However, more recently, repeated findings linking depression risk factors with relatively blunted cortisol reactivity have called this conceptualization into question (e.g., Oswald et al., 2006; Phillips, Carroll, Burns, & Drayson, 2005; Vrshek-Schallhorn et al., 2018).

Stress Induction in the Laboratory

Lab-based stress methodologies for examining stress-reactivity emerged in response to the need to study a consistent "dose" of stress in a controlled environment. For example, a widely used laboratory stress-induction paradigm, the Trier Social Stress Test (TSST; Kirschbaum, Pirke, & Hellhammer, 1993), consists of an extemporaneous speech and mental math calculation performed in the presence of an evaluative audience. Cortisol samples are collected prior to and several times after the stressor to measure the level of reactivity to the stressor. In a meta-analysis of 208 studies using the TSST and

other lab-based manipulations, Dickerson and Kemeny (2004) identified social evaluation and uncontrollability as statistically unique stressor characteristics predicting larger cortisol responses on average. Further, in this meta-analysis, a significant difference emerged between tasks with one versus two social evaluative components, with the latter producing a larger cortisol response, pointing to a dose-dependent response to social evaluation. Following the publication of that meta-analysis, several studies have sought to capitalize on findings regarding social evaluation by adapting the TSST to include explicitly negative non-verbal feedback (e.g., Taylor et al., 2010). Indeed, this so-called negative-evaluative TSST was shown to elicit a significantly larger cortisol response than a version of the TSST very similar to Kirschbaum et al.'s (1993) original stress paradigm (Vrshek-Schallhorn et al., 2018). These emerging findings suggest that social evaluation level is an important characteristic influencing cortisol reactivity and should be considered when examining depression risk-reactivity findings.

Trait Depression Risk Factors and Lab-based Stressor Reactivity

Studies of lab-based stress induction in adults have separately examined the influence of current depression and depression risk factors on cortisol reactivity; in both lines of research, evidence has been inconsistent regarding the direction of the link to depression. A meta-analysis of studies examining actively depressed individuals compared to non-depressed persons linked depression with a blunted cortisol response to a lab-based stressor (Burke, Davis, Otte, & Mohr, 2005). However, an interesting pattern was noted in this meta-analysis: More severe depression predicted blunted cortisol reactivity, while moderate depression predicted heightened cortisol reactivity. These

findings were empirically corroborated in a further study (Burke, Fernald, Gertler, & Adler, 2005), hinting that depression severity influences the direction of HPA axis dysregulation in clinical samples.

In non-clinical samples, studies examining the relationship between HPA axis functioning and depression risk factors in response to an acute stressor have also reported divergent findings. For example, neuroticism, low extraversion, and trait rumination were found to be associated with increased cortisol reactivity in response to a stressor (Wilson et al., 2015; Wirtz et al., 2007; Zoccola, Quas, & Yim, 2010). However, in separate studies, the opposite relationship emerged with the same risk factors predicting blunted cortisol reactivity to the stressor: neuroticism (Bibbey, Carroll, Roseboom, Phillips, & de Rooij, 2013; Phillips et al., 2005; Campos et al., 2014; Oswald et al., 2006), low extraversion (Bibbey et al., 2013; Oswald et al., 2006), and trait rumination (Vrshek-Schallhorn et al., 2018; Zoccola, Dickerson, & Zaldivar, 2008). An intriguing question is whether cross-study heterogeneity in social evaluation level (as a primary indicator of severity) might partially explain these divergent findings.

The Cortisol Reactivity Threshold Model: Cortisol as a Resource Mobilizing Hormone

The Cortisol Reactivity Threshold Model was developed in part based on an anecdotal observation that certain studies producing positive depression risk–reactivity associations used modest stressors, while others producing negative associations used more robust stressors (Vrshek-Schallhorn et al., 2018). This model suggests that individuals systematically differ in the level of stressor severity that elicits their "peak"

cortisol reactivity, such that those at elevated risk for depression generate peak cortisol reactivity at lower levels of stressor severity than do their lower risk counterparts. This prediction stems from two lines of evidence: Cortisol functions as a resource-mobilizing hormone, and individuals at elevated risk for depression perceive threat (and thus mobilize resources) at lower levels of stressor severity.

Several empirical findings support the perspective that cortisol functions as a resource-mobilizing hormone, which prepares the body to meet the demands of the day or the demands of an acute challenge (Fries et al., 2009; Wetherell, Lovell, & Smith, 2015). For example, a study examining affective and social characteristics and the CAR found that prior day feelings of being overwhelmed or threatened predicted an increased CAR the next day (Adam, Hawkley, Kudielka, & Cacioppo, 2006). In a study examining the relationship between the CAR and chronic stress, the CAR was lower on weekend days than on workdays, suggesting that more demanding days elicit a heightened CAR (Kunz-Ebrecht, Kirschbaum, Marmot, & Steptoe, 2004; Schlotz, Hellhammer, Schulz, & Stone, 2004). Additionally, individuals who were expecting to participate in a stressful task exhibited a higher CAR on the day of the challenge compared to a day without such challenge, suggesting that anticipation of a stressor mobilizes resources in preparation for response (Wetherell et al., 2015).

Cognitive theory (Beck, 1979) and evidence support that individuals at elevated risk for depression interpret there to be a threat with less provocation (i.e., a threat appraisal bias). A robust literature documents information processing biases to negative information in internalizing samples (for a review, see Gotlib & Joormann, 2010) and the

latent internalizing spectrum predicted exaggerated self-ratings of naturalistic stressor severity, as compared to investigator ratings (Conway, Starr, Espejo, Brennan, & Hammen, 2016).

Taken together, this suggests that heightened cortisol reactivity might represent engagement in a challenge, while blunted cortisol reactivity might represent a lack of engagement—but critically *either* due to feeling unperturbed or conversely overwhelmed by the challenge, leading to withdrawal. The Cortisol Reactivity Threshold Model suggests that the characteristics of the lab-based stress induction (chiefly the level of social evaluative threat) influence the range of responses that are most likely to be observed, and thus, influence how to interpret blunted responding. Specifically, modest stressors may be likely to elicit a relatively lower range of responses from unperturbed (blunted) to engaged (heightened) responses, while robust stressors may be likely to elicit a relatively higher range of responses from engaged (heightened) to overwhelmed (blunted) responses. The Cortisol Reactivity Threshold Model posits that individuals at risk for depression exhibit heightened cortisol responding to moderate stressors, whereas under robust stress they are unable to mount an appropriate response leading to a relatively blunted pattern compared to their lower-risk counterparts. Conversely, the model posits that those at lower risk for depression show an unperturbed or relatively blunted cortisol response to moderate stressors, which they perceive do not require resource mobilization, and they exhibit the expected and arguably adaptive elevated cortisol response to a robust stressor. This model predicts a distinct pattern of cortisol responding along a stressor severity continuum with at-risk individuals peaking earlier

than their low-risk counterparts, see Figure 1. Such a pattern of responding would generate positive risk-reactivity associations under moderate stressor severity and negative risk-reactivity associations under robust stressor severity, providing an explanation for the divergent findings in the literature. In a first test of this model, level of evaluative threat significantly interacted with the risk factor trait rumination to predict reactivity: under non-stressful conditions, no association was observed, while under modest threat a positive risk-reactivity relationship was observed, and under robust threat, a negative relationship was observed (Vrshek-Schallhorn et al., 2018).

The Present Study

The aim of this study was to address divergent findings in the depression risk factor and lab-based cortisol reactivity literature via a meta-analysis. More specifically, I evaluated the Cortisol Reactivity Threshold Model, which posits that the *severity* of the stressor (as indicated by degree of social evaluation) moderates the risk-reactivity relationship, such that moderate stressors produce positive risk-reactivity associations and robust stressors produce negative risk-reactivity associations. To accomplish this, I examined the relationship between trait-like depression risk factors (operationalized as extraversion, negative cognitive style, neuroticism, perfectionism, and trait rumination) and cortisol reactivity to a lab-induced psychosocial stressor. Additionally, I developed a system for assigning ratings for the severity of social evaluation in lab-induced stressors across the different studies for moderation analysis (see Method section). I hypothesized an interaction between social evaluation severity and depression risk factors, such that under moderate lab-based stressor severity, depression risk factors would be associated

with *heightened* cortisol reactivity to the stressor, while under robust stressor severity, depression risk factors would be associated with *blunted* cortisol reactivity. I did not predict a significant association under non-stressful conditions.

CHAPTER II

METHOD

Study Selection

Studies for this meta-analysis were selected based on five criteria: (1) participant characteristics, (2) depression risk factors, (3) psychological laboratory stressor tasks, (4) cortisol collection method, and 5) language. Studies that did not present new data, such as qualitative reviews and reanalysis of previously published data, were excluded.

Participant characteristics. Studies were included if they assessed adults (above 18 years of age) because cortisol response has been shown to change with developmental and pubertal status (Gunnar & Quevedo, 2007; Lupien, McEwen, Gunnar, & Heim, 2009). Studies examining clinical samples were included only if they also had a control sample, which was used in the meta-analysis. Clinical samples with an explicit depression diagnosis were excluded from the study based on evidence that severely depressed people exhibit a blunted response to a stressor (Burke et al., 2005).

Depression risk factors. A literature search was conducted in advance of the meta-analysis to identify trait-like risk factors for depression that have been studied in conjunction with cortisol reactivity; identified risk factors were included in the meta-analysis. For the purpose of this study, I defined trait-like depression risk factors as extraversion, negative cognitive style, neuroticism, perfectionism, and trait rumination, all of which have been shown to be stable trait-like characteristics associated with

depression. Despite mixed evidence regarding extraversion's link to depression, I chose to retain this risk factor to maximize study inclusion. I included all lab-based cortisol reactivity studies that examine at least one of these depression risk factors. Additionally, I included studies using constructs identified during the search that are substantially related to the included risk factors (i.e., harm avoidance as a construct related to neuroticism). Decisions regarding the suitability of a related construct were made in a consensus meeting. Studies examining other depression risk factors that are experiential, for example early childhood adversity, certain life events such as loss of a loved one, or that are not trait-like, such as state rumination or negative affect, were excluded.

Psychological laboratory stress task. This meta-analysis focused on psychological stressor tasks used to induce a cortisol response as defined by Dickerson and Kemeny's (2004) meta-analysis. These included: cognitive tasks (e.g., solving problems or performing mathematic calculations); public speech tasks that included verbal performance in front of others or verbal interaction with unknown others; and noise exposure and emotion induction tasks (e.g., watching or listening to emotional content) aimed at eliciting a negative mood state. Study conditions using overtly positive evaluations were excluded because of evidence that such positive appraisal also influences cortisol reactivity to lab-based stressors (Taylor et al., 2010). Additionally, studies using non-psychological forms of stress-induction such as physical stress exposure (e.g., cold-pressor tasks in which participants submerge a hand in ice water), a biological manipulation (e.g. dexamethasone suppression tasks), or naturalistic stress

exposure (taking an exam) were excluded from this meta-analysis because cortisol responses are thought to be stressor-specific (Kemeny, 2003).

Cortisol. Only studies assessing unbound salivary or plasma cortisol were included. Studies examining urinary cortisol were excluded because urinary cortisol reflects hormone levels accumulated over time rather than time-specific levels, which renders it inappropriate to assess short-term changes in cortisol levels (Kudielka, Gierens, Hellhammer, Wüst, & Schlotz, 2012).

Language. Only studies in English were examined to allow the authors to code the studies and to facilitate evaluation of inter-rater reliability.

Study identification. Psychinfo and PubMed databases were searched electronically with a combination of search terms used by Dickerson and Kemeny (2004) and terms unique to this meta-analysis across three categories: "*HPA, cortisol, hydrocortisone, psychoneuroendocrinology*" with, "*stress response, acute stress, laboratory stress, psychosocial stress, experimental stress,*" and "*neuroticism, extraversion, rumination, perfectionism negative cognitive style, cognitive vulnerability*." The search included all studies published by the end of the search date (September, 2018). Reference lists of included studies were examined to identify additional eligible studies. I also reviewed in full all studies from a relevant meta-analysis (n=186) examining stressor characteristics, which elicit a cortisol response (Dickerson & Kemeny, 2004). To identify unpublished literature, I or my mentor solicited unpublished studies examining the relationship between depression risk factors and cortisol response to an acute laboratory stressor from members of three academic societies through their email

listservs or social media: the International Society for Psychoneuroendocrinology (ISPNE), the Society for Research in Psychopathology (SRP), and the Association for Behavioral and Cognitive Therapies (ABCT). Additionally, I contacted authors who had multiple independently published eligible studies to inquire whether they had additional unpublished data. Finally, I or my mentor contacted several scholars known to conduct research in this area (depression risk factors and psychological laboratory controlled stress) that did not appear in the literature search.

Coded Variables

The author extracted the follow information from each study: (1) participant demographics, (2) depression risk factors, (3) cortisol characteristics, (4) stressor characteristics and (5) effect sizes characterizing the relationship between depression risk factors and cortisol reactivity. Additionally, a different individual blind to the risk-reactivity effect sizes coded de-identified study stressor characteristics to assign a stressor severity and uncontrollability ratings after which an interclass correlation coefficient was calculated to assess interrater reliability. I emailed study authors in situations requiring further stressor clarifications, and all discrepancies were addressed in a consensus meeting.

Participant characteristics. The following sample characteristics were extracted from each study: (1) sample size, (2) mean age and standard deviation, (3) gender composition, represented as a proportion (4) ethnic composition (non-white composition, represented as a proportion).

Depression risk factors. I extracted from each study type of depression risk factor (extraversion, negative cognitive style, neuroticism, perfectionism, and trait rumination) and coded it as a categorical variable. Effect sizes capturing depression risk as continuous or dichotomous variables (e.g., high vs. low neuroticism) were included and dichotomized (dimensional=0, categorical =1).

Cortisol characteristics. The following cortisol characteristics were extracted from each study: (1) cortisol source, (2) collection time, and (3) method of quantifying reactivity. The cortisol sample source produced a dichotomous variable (salivary = 0; plasma = 1). Cortisol collection time was recorded because evidence suggests a time of day effect for cortisol reactivity sampling (Burke et al., 2005; Dickerson & Kemeny, 2004). Basal cortisol levels are typically higher in the morning due to cortisol's diurnal secretion pattern, which can mask reactivity to a stressor. Collection time created a categorical variable indicating morning collection (before 12 PM, coded 0), afternoon collection (after 12 PM, coded 1), and both morning and afternoon collection time (coded 2). The method of quantifying cortisol reactivity (i.e. the type of data reduction) was recorded as a categorical variable, including: area under the curve with respect to increase (AUCi = 0); area under the curve with respect to ground (AUCg = 1); simple difference = 2, quadratic modeling = 3, and other = 4. See Table 1 for interpretation and derivation of cortisol reactivity indices.

Stressor characteristics. To provide an index of stressor severity, social evaluation characteristics were extracted from each stress task to yield a continuous variable from 0 (no social evaluation) to 2 (explicitly negative social evaluation).

No social evaluation (0 points). Stressors were coded as having no social evaluation under two conditions. First, if there was no evaluative audience present—for example the participant was completely alone, or an experimenter who was not providing any evaluation was present (e.g., experimenter is in the room but out of line of sight of participant and provides no feedback). The mere presence of such an individual is not associated with cortisol reactivity (Dickerson & Kemeny, 2004). Second, if a video camera or tape recorder was present, but participants were explicitly told that their performance will not be evaluated studies were coded as non-socially evaluative. Studies were excluded if participants had no audience but were video- or tape-recorded and told that their performance would be evaluated at a later time by experts or an audience; this creates some potential for social evaluation, but it is unclear based on the current literature how to best classify the severity of this manipulation.

Ambiguous social evaluation (1 point). I chose the term "ambiguous" to describe a range of neutral evaluation styles where participants could potentially infer negative evaluation, but where negative evaluation is not explicitly communicated through verbal or non-verbal behaviors. Stressors were coded as being ambiguous in social evaluation when there was at least one evaluative audience member present (other than the experimenter) whose purpose it was to view the participant's performance, or if the experimenter acted as an evaluative audience, for example by standing opposite and looking directly at the participant during the task. Empirical evidence suggests that to elicit a cortisol response an audience in addition to a non-evaluative experimenter must be present, as the presence of a non-evaluative spectator (e.g. experimenter is out of the

participant's line of sight or evaluative audience behind a one-way mirror or in another room) is not sufficient to elicit a cortisol response (Dickerson, Mycek, & Zaldivar, 2008). In order to be coded as ambiguous in social evaluation, protocols indicated that the audience member(s) displayed a neutral demeanor, which did not communicate an overtly negative or positive reaction to or evaluation of participants' performance (e.g., Kirschbaum et al.'s 1993 TSST).

Negative social evaluation (2 point). Stressors were coded as being negative in social evaluation when an audience was present and provided: 1) explicit negative evaluation such as overt nonverbal cues, which communicate dissatisfaction or boredom with performance (e.g., Vrshek-Schallhorn et al., 2018), or 2) misleading feedback of incompetent performance or harassment/critique during performance (e.g. "*You're obviously not good enough at doing this, now try harder. Keep going*!"; Earle, Linden, & Weinberg, 1999, p. 128). Although Dickerson and Kemeny (2004) include false feedback and audience criticism about performance as aspects of uncontrollability, a stressor characteristic they evaluate separately from social evaluation, it is reasonable to presume that participants are unaware of the uncontrollability of the manipulation and interpret this type of feedback as actual negative evaluation.

For studies reporting exact use of Kirschbaum et al.'s (1993) TSST as the stressor, stressor severity scores were calculated based on the original study to ensure consistency. In some studies it was unclear whether the exact Kirschbaum et al. (1993) standardized TSST stressor was used and whether modifications were made, and in these instances I contacted authors for clarification, and coded modifications relayed through

personal communication according to the characteristics described above. Similarly, when studies reported protocol modifications but did not specify the nature of the stressor adaptations, I contacted authors for clarification as needed. Final coded stressor severities are reported in Table 3.

Stressor severity index calculation. Stressor severity ratings were assigned as a minimum of 0 points for no social evaluation, 1 point for ambiguous social evaluation, and 2 points for negative social evaluation. However, the hypothesized influence of severity of social evaluation on the risk-reactivity relationship is not linear; rather, it follows a curvilinear relationship in which the risk-reactivity relationship is neutral under non-stressful conditions, positive under moderate severity, then negative under robust severity. This pattern can be modeled using a quadratic effect of the severity variable. Therefore, the linear variable (0, 1, 2) was squared to test the hypothesized quadratic relationship as follows: no social evaluation = 0 points; ambiguous social evaluation = 1 point; negative social evaluation = 4 points.

Coded variables related to stressor severity.

Uncontrollability. In coding uncontrollability, I followed Dickerson and Kemeny's (2004) coding guidelines. Studies were coded as uncontrollable if they included components, which could reasonably be expected to make participants feel they had limited control in influencing outcomes through their behavior or performance. The four stressor characteristics used by Dickerson and Kemeny (2004) were also used in this study to determined uncontrollable study aspects: (1) adaptions to tasks that made them impossible or very difficult to solve (2) deceptive negative evaluative feedback that the

participant is not performing well (e.g., indicating to participant that they have received fewer correct responses than they actually had); (3) audience censure about participant performance (e.g., telling participants to work faster); and (4) presentation of visual or auditory emotional content. Given findings that uncontrollability is a statistically unique predictor of greater cortisol response (Dickerson & Kemeny, 2004), I coded studies for uncontrollability despite my inclusion of some stressor characteristics such as false feedback and audience criticism about performance in the social evaluation variable and potential overlap between the two. Uncontrollability formed a dichotomous variable: (0=controllable; 1 uncontrollable).

Duration. Evidence suggests that stressor duration does not significantly predict the magnitude of cortisol response (Dickerson & Kemeny, 2004); however, anticipating a stressor has been shown to influence diurnal cortisol secretion (Wetherell et al., 2015), and may also influence cortisol reactivity. Thus, I have coded overall task duration, stressor duration, and stressor preparation period as continuous variables in minutes, and examined the influence of anticipation on cortisol reactivity specifically.

***Habituation*.** Studies in which the same participants underwent more than one stressor task of the same or similar type, or studies with different conditions conducted in the same environment (e.g., same lab) and same participants were coded for habituation. Though a previous meta-analysis used only data from a first stressor presentation to preclude a habituation effect (Dickerson & Kemeny, 2004), I included data from all repeated stressors from these studies and examined habituation (0 = only one presentation, 1 = multiple presentations) as a dichotomous moderator in analyses.

Evaluative audience size. Size of evaluative audience has been inconsistently related to cortisol reactivity. For example, one study reported a significantly higher cortisol response in a condition with 4 evaluative audience members compared to a condition with 1 audience member or a control group without an evaluative audience (0 versus 1 audience member did not differ significantly from one another; Bosch et al., 2009). In contrast, a further study found no significant difference in cortisol response in participants performing in front of one versus two audience members (Andrews et al., 2007). I extracted and coded audience size as a continuous variable (e.g., 0, 1, 2, or 3 audience members), and examined its potential influence on cortisol reactivity.

Statistics. For each study, I extracted the effect size for each individual risk factor-cortisol reactivity relationship, and reported multiple effects sizes for studies that provided these (e.g., a study reporting an association between neuroticism and AUCi as well as AUCg yielded two separate effect sizes). I recorded the type of statistic provided (zero-order Pearson correlation, partial correlation, t-statistic), as a categorical variable, and used this in moderation analyses.

Data Analysis

The Preferred Reporting Items for Systemic Reviews and Meta-Analyses (PRISMA) guidelines and flow diagram template were used to report total studies identified through the literature search, the process and justification for excluding studies, and final study count used in this meta-analysis (Liberati et al., 2009; Moher, Liberati, Tetzlaff, & Altman, 2009). I used Comprehensive Meta Analysis (CMA) Version 3 Software to conduct all analyses.

Hedges's *g* was used as the effect size to quantify the magnitude of the relationship between cortisol reactivity and depression risk factor across studies. It is well suited to this meta-analysis, which includes studies with small sample sizes because it is considered a conservative effective size measure for such small-sampled studies. Hedges's *g* was calculated from two different statistics, a regression model *t* statistic (predicting cortisol reactivity from risk factors), and Pearson correlation (*r* value relating risk factors with reactivity). In the absence of an appropriate effect size in the publication, I requested from authors either a zero-order correlation or *t*-statistic depending on their employed analyses. Studies that did not include the required effect size for the relationship between depression risk factor and cortisol reactivity are donated as "unreported effect size" in Table 3, and this information was obtained from the authors. Additionally, included data that has never been published (*n*=4) is denoted as unpublished in Table 3, and when possible the closest related paper is cited. In two instances, studies provided partial correlations, which were included in the meta-analysis. To calculate Hedges' *g*, first, *t* values were converted to a correlation by the CMA program using the following formula: *r*-value = $\sqrt{((t^2)/(t^2+df))}$. Next, all correlations were converted to Hedges' $g = \frac{r/\sqrt{(1-r^2)}}{\sqrt{(df(n_1+n_2)/n_1n_2)}}$

In two types of cases, it was necessary to reverse the direction of the reported effect sizes; see Table 3 for effect sizes as reported versus as used in analyses. First, the direction of an effect size for the relationship between extraversion and cortisol reactivity was switched (e.g., a positive zero order-correlation, *r*=0.3 was converted to a negative

zero-order correlation, r=-0.3) to capture the converse of extraversion, introversion, which is associated with increased risk for depression. If introversion rather than extraversion was reported, the sign was not reversed. Second, effect size signs were also changed for cortisol indices using quadratic time modeling because the inverted parabola capturing typical reactivity curves is defined by a negative sign, $y = -(x^2)$. In other words, a larger curve for cortisol reactivity yields a more negative *t*-value, while blunted cortisol reactivity with a flatter curve yields a *t*-value closer to zero.

In six instances, authors generously provided me with their raw data and I calculated the effect size between the depression risk factor and cortisol reactivity for inclusion in the meta-analysis. These studies are marked with a superscript "C" in Table 3. First, I calculated area under the curve with respect to increase (AUCi) in raw cortisol using a trapezoid formula (Pruessner, Kirschbaum, Meinlschmid, & Hellhammer, 2003), a common index of cortisol reactivity in the cortisol literature (Vrshek-Schallhorn et al., 2018; Zoccola et al., 2010). Second, AUCi values were examined for skew and natural log transformed (after adding a constant to make all values positive) if the skew statistic was greater than 1 or more than 3 times the standard error. In instances when log transformation did not appear to improve skew, I used untransformed AUCi values (n=2). Third, I calculated a zero-order correlation between AUCi and each depression risk factor available and used these *r*-values in the meta-analysis.

Multiple effect sizes reported in studies were all included but were treated according to whether they were independent versus dependent effect sizes. Studies that included multiple *independent* subgroups (e.g., findings for independent experimental

groups reported separately or results reported for males and females) were entered as separate studies (Borenstein, Hedges, Higgins, & Rothstein, 2009; Dickerson & Kemeny, 2004). For studies reporting multiple *dependent* effect sizes (e.g., using different reactivity metrics or trait-like risk factors), both effect sizes were entered as separate outcomes; however, these were pooled to create one unique risk-reactivity effect size for that study. Studies that did not provide appropriate statistics or indication of significance and direction of relationship were excluded after authors were contacted twice and did not respond.

Analyses used a random effects model as it allows for the variance to differ across studies (Borenstein et al., 2009). The random effects model assumes that variances across studies are due to both within-study differences, such as sampling error, *and* between-study variations, which reflect real difference, or heterogeneity between studies. By contrast, fixed effect models assume a common effect size across all studies and variations are attributed to random error, an inappropriate model for this meta-analysis, given the multiple characteristics that vary across studies (stressor, cortisol and participant characteristics).

I also conducted sensitivity analyses to examine the influence of individual studies on the overall effect size by systematically excluding one study at a time and recalculating the overall effect size. If a study is overly influential, the overall effect size will significantly differ once the study is removed from analyses.

I conducted moderation analyses, first to test my hypothesis that quadratic stressor severity significantly moderates the risk-reactivity relationship, and second to rule out

other simple explanations for divergent findings. I examined all continuous moderators, linear and quadratic stressor severity, age, gender, minority, audience size, and stressor preparation period, with meta-regression, an analysis analogous to linear regression with the effect size regressed onto the moderator variable. When examining stressor severity as a moderator, first, effect sizes utilizing all depression risk factors were examined together, and second, the individual depression risk factors were examined in separate analyses. This second set of analyses was run because some risk factors are more strongly linked to depression compared to others, which may impact meta-analytic findings. In all models examining stressor severity as a moderator, the linear and quadratic stressor severity terms were entered in the same meta-regression model together, analogous to the use of growth curve models to examine change over time (Hedeker & Gibbons, 2006). In moderation analyses for all other predictors, variables were entered in the regression model separately.

The following categorical variables, depression risk factor, uncontrollability, stressor habituation, timing of cortisol collection, cortisol source, cortisol reactivity index, and statistic type were examined in subgroup analyses, analogous to Analysis of Variance (ANOVA), with the exception that in a meta-analysis, between-study heterogeneity is divided into between and within study group variance. In the subgroup analyses, I pooled within study variance because it is reasonable to expect that there would not be variation in within-study dispersion across different groups (for example, depression risk factor or cortisol source). I only hypothesized that the quadratic stress term will moderate the risk-reactivity relationship, and did not anticipate moderating

effects of the remaining variables and thus made no hypotheses regarding their effects on the depression risk-cortisol reactivity relationship. In one instance, categorical versus dimensional treatment of depression risk factor, there were an insufficient number of studies to conduct moderation analyses (n=1 categorical study).

Moderators are study-level characteristics, meaning that they vary between and not within studies. This creates a dilemma when studies provide multiple dependent outcomes for a single moderator variable because effect sizes are pooled. In these instances, the pooled effect sizes are not interpretable in moderation analyses because they average across the different levels of the moderator (for example, combining AUCi and AUCg). Several studies included multiple outcomes for a specific moderator, for example, cortisol reactivity index, or type of effect size reported. Thus, these were dropped from moderation analyses (n=2) as the study cannot be assigned to a particular moderator group for analysis. It is important to note that in examining depression risk factor as a moderator, studies with multiple dependent effect sizes (e.g., extraversion and neuroticism) were treated as independent. While this violates the statistical independence assumption, it was necessary to decouple the effect sizes per depression risk factor, which were otherwise combined in all other analyses.

To test for heterogeneity of effect size across studies, I computed a Q statistic and an I^2 statistic. The Q statistic represents weighted squared deviations of individual effect sizes from the overall effect size, while I^2 provides a proportion of between study variation compared to total variation (Borenstein et al., 2009). When possible, I reported I^2, which is sensitive to study sample size (Higgins, Thompson, Deeks, & Altman, 2003).

I interpreted the magnitude of the I^2 statistics based on the Cochran Collaboration's suggested range guidelines: 0% to 40%, minimal level of heterogeneity; 30% to 60%, moderate heterogeneity; 50% to 90%, substantial heterogeneity, and 75% to 100%, considerable heterogeneity (Ryan, 2014).

Finally to test for publication bias, I calculated Egger's regression asymmetry test for bias using CMA, and examined the symmetry of a funnel plot. In Egger's linear regression the standardized effect size is regressed on the reverse standard error, or precision (Card, 2012). Funnel plots graphically depict the relationship between study sample size or variance and effect size. A funnel pattern indicates no bias when small studies show a pattern of equal distribution along the x-axis of large and small/non-significant effect size, while larger studies aggregate closer to the mean effect size. An asymmetrical funnel may indicate publication bias by depicting a lack of small studies without a significant effect size or systematic difference in effect sizes between smaller and larger studies.

CHAPTER III

RESULTS

Included Studies in Meta-Analysis

This meta-analysis included 29 studies, with 40 independent experimental samples for analyses. The initial database search produced 901 yields. An additional 1787 studies were identified through listserv/social media outreach, references, and direct author contacts resulting in a total of 2688 initial identified studies. Duplicate study removal resulted in 794 unique studies. After title and abstract screening, 363 articles remained to be reviewed in full. Studies were excluded based on pre-defined exclusion criteria (see Method section), and a final 29 studies, with 40 independent experimental samples, and 72 effect sizes were included in the meta-analysis, see Tables 2 and 3 for a list of included studies and Figure 2 for the flow diagram depicting the literature search process and final study inclusion. The 40 experimental samples were derived from the overall study count (*n*=29) some of which were entered as independent studies for the purpose of this meta-analysis, including studies providing two independent stress conditions (*n*=7), studies providing three independent stress conditions (*n*=1), and studies providing separate results for males and females (*n*=2). The 72 outcome effect sizes were derived from the 40 independent experimental samples as well as studies that provided multiple dependent outcomes including an effect size for two depression risk factors (*n*=7), three depression risk factors (*n*=2), eight depression risk factors (*n*=1), and two

cortisol reactivity indices (*n*=2). Of note, effect sizes for dependent outcomes within a single study were pooled and these studies provided one overall unique effect size for the meta-analysis.

Participant and Study Characteristics

Across the 40 independent experimental samples, a total of 1,994 individuals participated with an unweighted average mean age of 27.8 years (*SD*=3.81). Across the studies, an average of 47.6% of participants were female and the samples comprised on average 39.8% non-white participants, see Table 2 for study demographic characteristics.

Studies varied with regard to which depression risk factors they examined; most examined more than one. Twenty-four of the 40 independent experimental samples examined neuroticism and related constructs, followed by rumination (*n*=16), extraversion (*n*=13), perfectionism (*n*=6), and negative cognitive style (*n*=2).

There was also wide variability in the stress manipulations employed across studies, including experimental conditions examining no social evaluation (*n*=10), ambiguous social evaluation (*n*=27) and negative social evaluation (*n*=3), see Table 3 for study characteristics and statistics. Importantly, in coding stressor characteristics as defined by this meta-analysis, inconsistent language in the literature emerged describing the severity of social evaluation specifically in the ambiguous category. The level of evaluation was described as neutral in some instances (e.g., Kirschbaum et al., 1993; Campos et al., 2014) and negative (Zoccola & Dickerson 2015) or unfriendly in others (Ioannou, Furuya, & Altenmüller, 2016). In clarifying stressor characteristics through personal communication, it emerged that evaluation described as unfriendly or negative

did not always refer to the use of overt negative evaluation, rather it appears that the lack of positive evaluation created the *potential* for interpreting the provided feedback by the audience as negative. For example, in clarifying a "strict" behavior, C. I. Ioannou (personal communication, August 8, 2018) noted that the audience members immediately made a correction in performance mistakes but did not provide positive or negative feedback, while "unfriendly behavior" was characterized by facial expressions without either a positive or a negative tone. Thus, based on the absence of any overt negative verbal or nonverbal behaviors such studies were coded as ambiguous in evaluation for the purpose of this meta-analysis. For interrater reliability, the interclass correlation coefficient was 1.00 for stressor severity and 0.935 for uncontrollability.

Overall Effect Size

As anticipated, prior to accounting for stressor severity, the average overall effect size characterizing the relationship between combined depression risk factors and cortisol reactivity across all studies was not significant (g=0.039, p=0.609); see Figure 3a for effect sizes for each included experimental sample as well as the overall meta-analytic effect. The average overall effect sizes for the relationship between each separate depression risk factor and cortisol reactivity were all non-significant (extraversion, g=0.152, p=0.216; negative cognitive style g=-0.392, p=0.146; neuroticism, g=-0.059, p=0.527; perfectionism, g=0.075, p=0.623; rumination, g=0.077, p=0.604), see Figures 3b to 3f. These findings indicate that cortisol reactivity is not significantly associated with depression risk factors in aggregate. Analyses revealed that heterogeneity was present and significant for several groupings of studies, including studies with combined

depression risk factors (I^2 =55.813, p<0.001), extraversion (I^2=54.215, p=0.010), neuroticism (I^2=53.588, p=0.001), and rumination (I^2=67.096, p<0.001). The magnitude of I^2 reflects moderate to substantial heterogeneity as suggested by the Cochrane Collaboration guidelines, indicating real differences in study characteristics likely contributed to the variation in effect sizes. Heterogeneity was not significant for studies examining negative cognitive style, (I^2=41.148, p=0.192) or perfectionism (I^2=12.108, p=0.338) which may in part be due to the low number of studies for these risk factors; see Table 4 for all study statistics.

To probe the nature of this heterogeneity, moderation analyses were conducted. First, I examined the hypothesized moderator, quadratic stressor severity along with linear stressor severity. To rule out other simple explanations for the divergent findings, I followed up with further moderation analyses and tested the following variables: mean age, proportion female, proportion minority, audience size, and anticipation time (all continuous variables), habituation to stressor, cortisol source, time of cortisol collection, uncontrollability, depression risk factor, cortisol reactivity index, and statistic used (categorical variables). All moderation analyses were conducted across the combined depression risk factors to maximize statistical power with the exception of analyses examining the role of stressor severity on depression risk factors, as described previously. Additionally, in two depression risk factors, perfectionism and negative cognitive style, I was unable to examine the role of stressor severity. First, both perfectionism and negative cognitive style had only two levels of the moderator (no social evaluation and ambiguous social evaluation, lacked negative social evaluation), which precluded conducting

moderation analysis across the three levels. Second, negative cognitive style had only two experimental samples, with the minimum required being three, which precluded further analyses.

Sensitivity Analyses

I conducted sensitivity analyses to examine whether any one study overly influenced the overall effect size and results indicate no one study was unduly influential after exclusion of each study and recalculation of an overall effect size, all *ps*>0.05; see Figure 4.

Continuous Predictors of the Risk-Reactivity Effect Size

Moderation analyses were conducted with meta-regression for continuous variables. Quadratic stressor severity, my hypothesized moderator, was not a significant predictor of the risk-reactivity association for the combined depression risk factors (β=-0.251, p=0.163), extraversion (β=-0.090, p=0.751), or neuroticism (β=-0.321, p=0.118); however, it approached significance for rumination (β=-0.593, p=0.062). I did not further decompose this quadratic effect by condition because it was not statistically significant; the effect size direction, however, is consistent with an inverted parabola, with relatively higher values in the center as compared to the extremes. As anticipated, linear stressor severity did not significantly predict the risk-reactivity effect size for combined depression risk factors (β=0.254, p=0.431), extraversion, (β=0.173, p=0.768), neuroticism, (β=0.207, p=0.570), or rumination (β=0.888, p=0.149).

Next, other continuous moderators were examined. Proportion female of the sample significantly predicted the risk-reactivity effect size (β=-0.598, p=0.038), such

that a greater proportion of females in a sample was associated with a more negative risk-reactivity relationship. The length of the anticipation phase also significantly predicted the risk-reactivity relationship (β=-0.069, p=0.003) in the same direction, such that increased anticipation time prior to a stressor was associated with a more negative risk-reactivity relationship. Given the significant finding regarding anticipation, I examined whether anticipation uniquely predicted the risk-reactivity effect size, or whether longer tasks in general predict a negative risk-reactivity association. I conducted two more meta-regression analyses with post-anticipation task time and task time plus anticipation time combined and neither was significant, (β=0.021, p=0.154; β=-0.003, p=0.844, respectively), indicating the effect of anticipation duration was not driven by longer tasks post-anticipation. All other examined predictors were not significant, specifically proportion minority (β=-0.074, p=0.821), audience size (β=0.030, p=0.713) and mean age (β=0.003, p=0.620), see Table 5 for all statistics.

Subgroup Analyses of Categorical Moderators

Subgroup moderation analyses were conducted for categorical variables. Although I had made no predictions regarding habituation, results indicated a significant difference in effect size between studies using habituation compared to those without habituation (Q=7.455, p=0.006). Specifically, studies with habituation produced a significant, positive risk-reactivity association (g=0.541, p=0.007), whereas those without habitation produced a non-significant effect size (g=-0.043, p=0.583). There was no evidence that effect sizes differed significantly for any of the remaining categorical variables. The risk-reactivity effect size was not significantly different between studies,

which examined extraversion, negative cognitive style, neuroticism, perfectionism, or rumination (Q=4.646, p=0.326; all individual effect sizes between g=-0.392 and 0.169, all ps>0.05). There were also no significant group differences in effect size between studies using salivary versus plasma cortisol (Q=0.591, p=0.442; salivary, g=0.057, p=0.481; plasma, g=-0.169, p=0.549). Similarly, the effect size did not differ significantly across cortisol reactivity indices (Q=7.376, p=0.117, all gs between, -0.713 and 0.345, other p=0.051, all other ps>0.05). No significant difference in the risk-reactivity effect size emerged between studies collecting cortisol in the morning, afternoon, or morning and afternoon, (Q=1.042, p=0.594; gs between 0.018 and 0.333, all ps >0.05). Similarly, studies differing in the uncontrollability of stressors also produced non-significant results (Q=0.090, p=0.764 controllable, g=-0.021, p=0.924, uncontrollable g=0.049, p=0.555). Finally, statistic type used did not produce significant differences in effect size across the different groups (Q=1.545, p=0.462, gs between 0.031 and 0.642, all ps>0.05). See Table 6 for all statistics.

Publication Bias

To examine publication bias in this meta-analysis, I first examined a funnel plot, which visually depicts the distribution of the effect sizes on the x-axis and standard error on the y-axis. In the absence of publication bias I would expect a fairly even distribution of studies around the mean. The funnel plot, Figure 5, shows that studies are fairly symmetrically distributed indicating no obvious publication bias. I also conducted an Egger's regression asymmetry test for publication bias, and results again indicated no publication bias (t=1.92, SE=0.556, p=0.061).

CHAPTER IV

DISCUSSION

In this meta-analysis, I examined the association between trait-like depression risk factors and cortisol reactivity in 40 independent experimental samples and tested the hypothesis that quadratic stressor severity would moderate the risk-reactivity relationship as posited by the Cortisol Reactivity Threshold Model. Findings indicated no significant overall effect size between cortisol reactivity and depression risk factors, combined or separately examined as anticipated. Contrary to my hypothesis, quadratic stressor severity did not significantly predict the risk-reactivity association in combined depression risk factors, extraversion or neuroticism; however, in rumination, quadratic stress approached significance in predicting that relationship. My confidence in these null findings is tempered by the few number of negative evaluative studies available for this meta-analysis, as well as by my observation that among studies using ambiguous stressors, severity was heterogeneous in a way that was not always well documented in methods sections or possible to capture with my three-level severity indicator. Several unpredicted findings also emerged. Both proportion female and stressor anticipation emerged as significantly predicting relatively blunted risk-reactivity relationships (higher risk, lower reactivity). Additionally, a significant difference in risk-reactivity effect size emerged between studies using habituation versus those that did not, with studies using habituation yielding a significant and positive risk-reactivity relationship.

No other continuous dichotomous moderator variables significantly related to risk-reactivity effect size.

Overall my findings did not provide supporting evidence for the role of stressor severity in the risk-reactivity association as predicted by the Cortisol Reactivity Threshold Model; however, some important considerations are warranted. First, the null findings for the separate depression risk factors (neuroticism, rumination, and extraversion) were impacted by the limited number of negative evaluative studies (n=3) included. An initial survey of the literature suggested that more such studies would be available. However, as noted in the results section, there was wide heterogeneity in the language used to describe ambiguous evaluative stressors, with some studies describing them as neutral as in the original Kirschbaum et al. (1993) TSST, and others describing this type of evaluation valence as unfriendly (Ioannou et al., 2016). While the obvious lack of any positive evaluation (e.g., neutral facial expression, immediate corrective feedback) could be considered negative by some, these stressors appear to be more ambiguous or neutral in tone in the absence of any explicit negative verbal or nonverbal behaviors. Quite understandably, given the *potential* for participants to interpret this as threatening, these evaluations have been described as negative evaluative in some studies. It may be, however, that this represents not just a semantic difference but rather a meaningful difference between studies in the severity of the "delivered" stressor that it is not possible to capture according to the a priori stressor severity definitions in the present meta-analysis. Thus, several studies that initially appeared negative evaluative were ultimately categorized as ambiguous limiting the number of negative evaluative studies to

three in the meta-analysis; hence results involving severity should be considered preliminary.

Second, no negative evaluative studies were available for either perfectionism (total experimental samples = 6) or for negative cognitive style (total experimental samples =2). Thus, no conclusions can be drawn for the influence of severity on risk-reactivity relationships for these risk factors.

Third, for one of the depression risk factors examined, extraversion, there is mixed evidence regarding its relationship to depression risk. One meta-analysis provided strong evidence for low extraversion being associated with depression (Malouff et al., 2005), while another meta-analysis found mixed and inconsistence evidence of extraversion's relationship to depressive disorders (Kotov et al., 2010). It appears that depression risk factors fall on a continuum such that some are more strongly associated with depression, such as neuroticism and rumination, while others are more weakly associated with the risk, such as extraversion. It is possible that as depression risk diminishes, it no longer shows the same pattern of relationship to cortisol reactivity. This could lead to results being washed out when depression risk factors are combined and could contribute to null findings in the combined risk factor group.

Fourth, with the exception of two studies, all correlational statistics used in analyses were zero-order correlations; these were preferred over partial correlations for uniformity and to provide a relatively conservative estimate of effects. Some evidence, however, suggests that not covarying important cortisol-related covariates biases effect sizes toward zero, which could have impacted meta-analytic results (Kudielka, Broderick,

& Kirschbaum, 2003). An intriguing future direction would be to examine the findings when partialing out covariates.

Several unpredicted findings also emerged in analyses. First, stress anticipation duration emerged as a significant predictor of a negative risk-reactivity relationship. Stress anticipation length may heighten stressor severity as it is reasonable to expect that as anticipation time prior to a stressor increases, stressor severity is perceived as more threatening, or severe. In support of this, anticipation of a natural stressor has been previously associated with an increased cortisol response prior to the stressor (Smyth et al., 1998). In a further study, participants were assigned to one of two conditions, anticipation-only (informed about upcoming task but after 10 minute anticipation time informed there would actually be no stressor) and stressor-only task (informed about task and engaged in stressor immediately thereafter, Engert et al., 2013). Results showed that even in the anticipation-only condition, participants exhibited increased cortisol levels compared to baseline levels, which could be consistent with an interpretation that increasing anticipation heightens perception of stressor threat. This finding thus could be viewed as consistent with the Cortisol Reactivity Threshold Model, which predicts a more negative risk-reactivity association under more severe stressors.

A second unpredicted finding emerged with regard to each sample's proportion of female participants, such that a higher proportion of female participants was associated with a negative risk-reactivity association. Indeed, that female gender predicts greater risk for depression is one of the most consistently reported findings in psychopathology research (Nolen-Hoeksema, Larson, & Grayson, 1999) Thus, similar to the effect of

anticipation duration, this finding could be interpreted as consistent with the Cortisol Reactivity Threshold Model, where females will be shifted towards more risk and blunted cortisol response (a shift left on Figure 1). Though this meta-analysis was not focused on examining the influence of gender on findings, I would expect that milder studies may show a positive risk-reactivity association, and future studies should investigate this potential.

The third unpredicted finding was related to habituation through repeated presentation of similar stressors. A significant difference between studies with habituation compared to those without emerged; specifically, studies that used habituation yielded a significant positive risk-reactivity effect size. This suggests that as individuals become more accustomed to a stressor they respond with heightened cortisol reactivity because the stressor is less novel and therefore less severe. These findings could also be interpreted as consistent with the Cortisol Reactivity Threshold Model, as one would expect less severe stressors to yield a positive risk-reactivity association.

Finally, the non-significant finding for uncontrollability merits discussion. In their seminal meta-analysis of stressor characteristics that elicit a cortisol response, Dickerson and Kemeny (2004), found that uncontrollability was a unique predictor in provoking cortisol reactivity. First, it is important to point out that the Dickerson and Kemeny's (2004) meta-analyses examined cortisol *reactivity*, while the present study examined the relationship *between* cortisol reactivity and depression risk factors; thus, the two findings are not directly comparable and this meta-analysis does not represent non-replication of the original finding. Second, Dickerson and Kemeny (2004) included in uncontrollability

aspects that I captured in social evaluation, such as receiving false negative feedback or criticism from the audience. Future research could examine the role of uncontrollably in the risk-reactivity relationship if social evaluative components such as false feedback and criticism are removed from the variable, but this is beyond the scope of this meta-analysis.

Future Directions

Despite the overall negative meta-regression results for the primary hypothesis, a quadratic effect of severity on the risk-reactivity relationship, several recommendations for future investigations in this area are merited. First, the field would benefit from conducting additional explicitly negative evaluative lab-based stress studies, as quadratic effect results approached significance for one risk factor, rumination, suggesting this may be a promising methodology. Second, I was unable to examine stressor severity in two depression risk factors, perfectionism and negative cognitive style; researchers should consider including these along with other trait-like depression risk factors when examining stressor severity.

Third, although evidence that stressor severity is important is still preliminary, based on the observed heterogeneity of methods within the ambiguous stressors, the field would benefit from further enhancing the standardization of stressor protocols. Standardization of stressors would likely aid in replication of findings and would also systematize language used in the field when describing stressor severity, in particular the level of social evaluation. Much of the heterogeneity in study descriptions appears to be due to how the *potential* for negative evaluation by an audience is described in

publications, with some studies describing what I have called "ambiguous" severity as neutral and others, understandably, as negative. One way to achieve greater standardization would be for all studies to report in the Method section: (1) the *intended* explicit type of evaluation by an audience (e.g., ambiguous/neutral evaluative, explicitly negative evaluative, no-audience non-evaluative control, explicitly positively evaluative), and (2) provide examples of the verbal and non-verbal behaviors used by the audience in support of the intended explicit evaluation type. Table 7 provides examples of language for each stressor level. A third possible future direction is for participants to provide subjective ratings of perceived evaluation (overall level, positive, and negative) and possibly other dimensions (perceived difficulty, challenge) to characterize the effect of the manipulation for comparison across studies and labs and to corroborate that the stressor has achieved its desired goal. Additionally, this would shed light on whether indeed there is a perceived difference between ambiguous and negative evaluative stressors.

Limitations

Despite several strengths including extensive gray literature searches, use of unpublished data, extensive personal communication with authors to clarify stressor ambiguities and other study details, and blind coding of stressor severity, this meta-analysis is not without limitations. First, as noted above, I was only able to include a limited number of negative evaluative studies (n=3), which reduced power substantially to detect an effect size and also precluded analyses in all depression risk factor groups. Further research can address this gap. Second, I excluded several studies whose stressors

I was unable to code for social evaluation based on my coding criteria. In this group of excluded studies, an audience was not directly present during the presentation of the stressor but participants were told that they were being evaluated (i.e., manipulation included a real or ostensible audience behind a one-way mirror or in a separate room, or the prospect of future evaluation of the video by judges). These studies were not well suited for inclusion in the ambiguous evaluative or negative evaluative condition based on prior evidence that the presence of an evaluative member is needed to elicit a cortisol response (Dickerson & Kemeny, 2004). They were also not appropriate candidates for inclusion in the non-evaluative group because there was *some* level of evaluation communicated to participants. While I attempted to capture different levels of severity in this study, it is possible that my 3-level scale is not fine-grained enough for some of the nuanced differences across studies. An intriguing question is whether fine-tuning my stressor severity measure to include some of these "in-between" conditions (using one-way mirrors, purported evaluation of videos, future evaluation) might produce different results. A third limitation was my exclusion of some studies that did not provide sufficient information for stressor severity to be coded or did not provide a risk-reactivity effect size. I contacted authors twice before excluding studies in the case of no response; however, this did reduce the final sample size. A fourth limitation is the combination across all depression risk factors in moderation analyses, as they may vary in potency. While I decided on this approach to maximize power in moderation analyses, additional research across different risk factors would allow for these risk variables to be examined independently in moderation analyses. A fifth limitation is my combination of non-

stressful and stressful conditions in analyses. Per the Cortisol Reactivity Threshold Model, the risk-reactivity relationship is moderated by stressor severity, thus, it was important to code for and confirm that no significant risk-reactivity association was observed in this non-stressful condition, and indeed I made no predictions about significant findings in this group. However, a non-significant effect in the non-stressful condition could potentially diminish possible overall significant effect findings across the two stressful conditions when combined or in analyses, which did not consider stressor severity. A sixth and final limitation was the unaccounted-for variability in the samples, including health variables such as use of oral contraceptives. While this study did not focus on examining different health variables, which may influence cortisol reactivity, it is possible that these factors did in fact influence results. Specifically, some research has shown that oral contraceptives are associated with a blunted cortisol responding to an acute stressor (Kirschbaum, Pirke, & Hellhammer, 1995). Future empirical studies should examine the role of oral contraceptives and other health variables as they relate to risk-reactivity findings.

Conclusions

Overall, I was unable to provide meta-analytic supporting evidence for the Cortisol Reactivity Threshold Model and the role of stressor severity in explaining the divergent risk-reactivity findings in the literature, although a strong test of this hypothesis was hampered by the low number of available negative-evaluative studies and the possibility of methodological heterogeneity within the largest severity group, ambiguous/neutral severity. However, some unpredicted findings emerged, which could

be viewed as consistent with the model. Together with a quadratic effect of severity approaching significance for the rumination risk-reactivity relationship, these findings suggest that further investigation of the Cortisol Reactivity Threshold Model is merited. Two of the findings are particularly intriguing because they indirectly tap into stressor severity. Longer anticipation of a stressor, which is thought to increase stressor severity, predicted a blunted risk-reactivity relationship while habituation, which would decrease the stressor severity, was associated with a positive risk-reactivity association, both model-consistent findings. Given the general lack of studies using more severe stressors and the general heterogeneity of studies using psychosocial stressors, further research, and greater stressor standardization are needed to provide more conclusive evidence for the role of stressor severity on the risk-reactivity relationship.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Abramson, L. Y., Metalsky, G. I., & Alloy, L. B. (1989). Hopelessness depression: A theory-based subtype of depression. *Psychological review*, *96*(2), 358-372.

Adam, E. K., Hawkley, L. C., Kudielka, B. M., & Cacioppo, J. T. (2006). Day-to-day dynamics of experience–cortisol associations in a population-based sample of older adults. *Proceedings of the National Academy of Sciences, 103*(45), 17058-17063.

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217-237.

Alloy, L. B., Abramson, L. Y., Hogan, M. E., Whitehouse, W. G., Rose, D. T., Robinson, M. S., ... & Lapkin, J. B. (2000). The Temple-Wisconsin cognitive vulnerability to depression project: Lifetime history of Axis I psychopathology in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology, 109*(3), 403-418

Alloy, L. B., Abramson, L. Y., Keyser, J., Gerstein, R. K., & Sylvia, L. G. (2008). Negative cognitive style. In K. S. Dobson & D. J. A Dozois (Eds.), *Risk Factors in Depression*, (pp.237-262). https://doi.org/10.1016/B978-0-08-045078-0.X0001-9

Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., Hogan, M. E., Panzarella, C., & Rose, D. T. (2006). Prospective incidence of first onsets and recurrences of depression in individuals at high and low cognitive risk for depression. *Journal of Abnormal Psychology, 115*(1), 145-156.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Andrews, J., Wadiwalla, M., Juster, R. P., Lord, C., Lupien, S. J., & Pruessner, J. C. (2007). Effects of manipulating the amount of social-evaluative threat on the cortisol stress response in young healthy men. *Behavioral Neuroscience, 121*(5), 871-876.

Barlow, D. H., Sauer-Zavala, S., Carl, J. R., Bullis, J. R., & Ellard, K. K. (2014). The nature, diagnosis, and treatment of neuroticism: Back to the future. *Clinical Psychological Science, 2*(3), 344-365.

Beck, A.T., Rush, A.J., Shaw, B.F., & Emery, G. (1979). *Cognitive therapy of depression*. New York: The Guilford Press.

*Bibbey, A., Carroll, D., Roseboom, T. J., Phillips, A. C., & de Rooij, S. R. (2013). Personality and physiological reactions to acute psychological stress. *International Journal of Psychophysiology, 90*(1), 28-36.

Borenstein, B., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to Meta-Analysis*. United Kingdom: John Wiley & Sons Ltd.

Bosch, J. A., De Geus, E. J., Carroll, D., Goedhart, A. D., Anane, L. A., van Zanten, J. J. V., ... & Edwards, K. M. (2009). A general enhancement of autonomic and

cortisol responses during social evaluative threat. *Psychosomatic Medicine*, *71*(8), 877-885.

Boyce, P., Parker, G., Barnett, B., Cooney, M., & Smith, F. (1991). Personality as a vulnerability factor to depression. *The British Journal of Psychiatry, 159*(1), 106-114.

Burke, H. M., Davis, M. C., Otte, C., & Mohr, D. C. (2005). Depression and cortisol responses to psychological stress: A meta- analysis. *Psychoneuroendocrinology, 30*(9), 846-856.

Burke, H. M., Fernald, L. C., Gertler, P. J., & Adler, N. E. (2005). Depressive symptoms are associated with blunted cortisol stress responses in very low-income women. *Psychosomatic Medicine, 67*(2), 211-216.

*Calingaert, N. (2016). The Role of Emotion Dysregulation in Atypical Cortisol Responses to Acute Psychosocial Stress in Individuals High in Neuroticism (Doctoral dissertation, Roosevelt University).

*Campos, B., Busse, D., Yim, I. S., Dayan, A., Chevez, L., & Schoebi, D. (2014). Are the costs of neuroticism inevitable? Evidence of attenuated effects in US Latinas. *Cultural Diversity and Ethnic Minority Psychology, 20*(3), 430-440.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: The Guilford Press.

Clark, L. A., Watson, D., & Mineka, S. (1994). Temperament, personality, and the mood and anxiety disorders. *Journal of Abnormal Psychology, 103*(1), 103-116.

Conway, C. C., Starr, L. R., Espejo, E. P., Brennan, P. A., & Hammen, C. (2016). Stress responsivity and the structure of common mental disorders: Transdiagnostic internalizing and externalizing dimensions are associated with contrasting stress appraisal biases. *Journal of Abnormal Psychology, 125*(8), 1079-1089.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment, 4*(1), 5-13.

Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin, 130*(3), 355-391.

Dickerson, S. S., Mycek, P. J., & Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychology, 27*(1), 116-121.

*Dinzeo, T. J., Sledjeski, E., Durner, C., & Docherty, N. M. (2015). Comparative study of personality trait characteristics and reactivity in schizophrenia using a film clip paradigm. *The American Journal of Psychology, 128*(4), 515-526.

Earle, T. L., Linden, W., & Weinberg, J. (1999). Differential effects of harassment on cardiovascular and salivary cortisol stress reactivity and recovery in women and men. *Journal of Psychosomatic Research, 46*(2), 125-141.

Engert, V., Efanov, S. I., Duchesne, A., Vogel, S., Corbo, V., & Pruessner, J. C. (2013). Differentiating anticipatory from reactive cortisol responses to psychosocial stress. *Psychoneuroendocrinology, 38*(8), 1328-1337.

Fanous, A. H., Neale, M. C., Aggen, S. H., & Kendler, K. S. (2007). A longitudinal study of personality and major depression in a population-based sample of male twins. *Psychological Medicine, 37*(8), 1163-1172.

Ferrari, A. J., Charlson, F. J., Norman, R. E., Patten, S. B., Freedman, G., Murray, C. J., ... & Whiteford, H. A. (2013). Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Medicine, 10*(11), e1001547.

Fries, E., Dettenborn, L., & Kirschbaum, C. (2009). The cortisol awakening response (CAR): Facts and future directions. *International Journal of Psychophysiology, 72*(1), 67-73.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research, 14*(5), 449-468.

*Gerra, G., Zaimovic, A., Franchini, D., Palladino, M., Giucastro, G., Reali, N., ... & Brambilla, F. (1998). Neuroendocrine responses of healthy volunteers to techno-music: Relationships with personality traits and emotional state. *International Journal of Psychophysiology, 28*(1), 99-111.

*Gerra, G., Zaimovic, A., Mascetti, G. G., Gardini, S., Zambelli, U., Timpano, M., ... & Brambilla, F. (2001). Neuroendocrine responses to experimentally-induced psychological stress in healthy humans. *Psychoneuroendocrinology, 26*(1), 91-107.

*Gianferante, D., Thoma, M. V., Hanlin, L., Chen, X., Breines, J. G., Zoccola, P. M., & Rohleder, N. (2014). Post-stress rumination predicts HPA axis responses to repeated acute stress. *Psychoneuroendocrinology, 49*, 244-252.

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology*, *6*(1), 285-312.

Gunnar, M., & Quevedo, K. (2007). The neurobiology of stress and development. *Annual Review of Psychology, 58*(1), 145-173.

Hammen, C. (2006). Stress generation in depression: Reflections on origins, research, and future directions. *Journal of Clinical Psychology*, *62*(9), 1065-1082.

Hammen, C., Davila, J., Brown, G., Ellicott, A., & Gitlin, M. (1992). Psychiatric history and stress: Predictors of severity of unipolar depression. *Journal of Abnormal Psychology*, *101*(1), 45-52.

Hankin, B.L, Abramson, L.Y., Miller, N. & Haeffel, G.J. (2004). Cognitive vulnerability-stress theories of depression: Examining affective specificity in the prediction of depression versus anxiety in three prospective studies. *Cognitive Therapy and Research*, *28*(3) 309-345.

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis* (Vol. 451). John Wiley & Sons.

Heim, C., Newport, D. J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R.,...Nemeroff, C. B. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *JAMA*, *284*(5), 592-597.

*Henckens, M. J., Klumpers, F., Everaerd, D., Kooijman, S. C., van Wingen, G. A., & Fernández, G. (2016). Interindividual differences in stress sensitivity: Basal and stress-induced cortisol levels differentially predict neural vigilance processing under stress. *Social Cognitive and Affective Neuroscience, 11*(4), 663-673.

Hewitt, P. L., Flett, G. L., & Ediger, E. (1996). Perfectionism and depression: Longitudinal assessment of a specific vulnerability hypothesis. *Journal of Abnormal Psychology, 105*(2), 276-280.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *The British Medical Journal*, 327(7417) 557-560.

Iacoviello, B. M., Alloy, L. B., Abramson, L. Y., Whitehouse, W. G., & Hogan, M. E. (2006). The course of depression in individuals at high and low cognitive risk for depression: A prospective study. *Journal of Affective Disorders*, *93*(1–3), 61-69.

*Ioannou, C. I., Furuya, S., & Altenmüller, E. (2016). The impact of stress on motor performance in skilled musicians suffering from focal dystonia: Physiological and psychological characteristics. *Neuropsychologia, 85*, 226-236.

Kemeny, M. E. (2003). The psychobiology of stress. *Current Directions in Psychological Science*, *12*(4), 124-129.

Kendall, A. D., Zinbarg, R. E., Mineka, S., Bobova, L., Prenoveau, J. M., Revelle, W., & Craske, M. G. (2015). Prospective associations of low positive emotionality with first onsets of depressive and anxiety disorders: Results from a 10-wave latent trait-state modeling study. *Journal of Abnormal Psychology*, *124*(4), 933-943.

Kendler, K. S., Karkowski, L. M., & Prescott, C. A. (1999). Causal relationship between stressful life events and the onset of major depression. *American Journal of Psychiatry*, *156*(6), 837-841.

Kendler, K. S., Kuhn, J., & Prescott, C. A. (2004). The interrelationship of neuroticism, sex, and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry*, *161*(4), 631-636.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heathe, A.C., & Eaves, L.J. (1993). A longitudinal twin study of personality and major depression in women. *Archive of General Psychiatry*, *50*(11), 853-862.

Kirschbaum, C., Pirke, K., & Hellhammer, D.H. (1993). The 'Trier Social Stress Test': A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*(1-2), 76-81.

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1995). Preliminary evidence for reduced cortisol responsivity to psychological stress in women using oral contraceptive medication. Psychoneuroendocrinology, *20*(5), 509-514.

*Kirschbaum, C., Prussner, J. C., Stone, A. A., Federenko, I., Gaab, J., Lintz, D., ... & Hellhammer, D. H. (1995). Persistent high cortisol responses to repeated psychological stress in a subpopulation of healthy men. *Psychosomatic Medicine*, *57*(5), 468-474.

Kotov, R., Gamez, W., Schmidt, F., & Watson, D. (2010). Linking "big" personality traits to anxiety, depressive, and substance use disorders: A meta-analysis. *Psychological Bulletin*, *136*(5), 768-821.

Kudielka, B.M., Broderick, J. E., & Kirschbaum, C. (2003). Compliance with saliva sampling protocols: Electronic monitoring reveals invalid cortisol profiles in noncompliant subjects. *Psychosomatic Medicine, 65*(2), 313-319.

Kudielka, B. M., Gierens, A., Hellhammer, D. H., Wüst, S., & Schlotz, W. (2012). Salivary cortisol in ambulatory assessment—some dos, some don'ts, and some open questions. *Psychosomatic Medicine, 74*(4), 418-431.

Kunz-Ebrecht, S. R., Kirschbaum, C., Marmot, M., & Steptoe, A. (2004). Differences in cortisol awakening response on work days and weekends in women and men from the Whitehall II cohort. *Psychoneuroendocrinology, 29*(4), 516–528.

*Lewis, E. J., Yoon, K. L., & Joormann, J. (2018). Emotion regulation and biological stress responding: Associations with worry, rumination, and reappraisal. *Cognition and Emotion, 32*(7) 1487-1498.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLOS Medicine, 6*(7), e1000100.

Lupien, S. J., McEwen, B. S., Gunnar, M. R., & Heim, C. (2009). Effects of stress throughout the lifespan on the brain, behaviour and cognition. *Nature Reviews Neuroscience, 10*(6), 434-445.

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment, 27*(2), 101-114.

*Marin, M. F., Pilgrim, K., & Lupien, S. J. (2010). [Modulatory effects of stress on reactivated emotional memories]. Unpublished raw data.

*McGirr, A., & Turecki, G. (2009). Self-critical perfectionism is associated with increases in sympathetic indicators in a controlled laboratory stress paradigm. *Psychosomatic Medicine, 71(5), 588.*

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264-269.

Monroe, S. M. & Simons, A. D. (1991). Diathesis—stress theories in the context of life stress research: Implications for the depressive disorders. *Psychological Bulletin, 110*(3), 406-425.

*Morris, M. C., Rao, U., Wang, L., & Garber, J. (2014). [Cortisol reactivity to experimentally manipulated psychosocial stress in young adults at varied risk for depression]. Unpublished raw data.

Naragon-Gainey, K., Watson, D., & Markon, K. E. (2009). Differential relations of depression and social anxiety symptoms to the facets of extraversion/positive emotionality. *Journal of Abnormal Psychology*, *118*(2), 299-310.

Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depression episodes. *Journal of Abnormal Psychology, 100*(4), 569-582.

Nolen-Hoeksema, S. (2000). The role of rumination in depressive disorders and mixed anxiety/depressive symptoms. *Journal of Abnormal Psychology*, *109*(3), 504-511.

Nolen-Hoeksema, S., Larson, J., & Grayson, C. (1999). Explaining the gender difference in depressive symptoms. *Journal of Personality and Social Psychology, 77*(5), 1061-1072.

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400-424.

*Oswald, L. M., Zandi, P., Nestadt, G., Potash, J. B., Kalaydjian, A. E., & Wand, G. S. (2006). Relationship between cortisol responses to stress and personality. *Neuropsychopharmacology, 31*(7),1583-1591.

*Phillips, A. C., Carroll, D., Burns, V. E., & Drayson, M. (2005). Neuroticism, cortisol reactivity, and antibody response to vaccination. *Psychophysiology*, *42*(2), 232-238.

Pruessner, J. C., Kirschbaum, C., Meinlschmid, G., & Hellhammer, D. H. (2003). Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology, 28*(7), 916-931.

*Puig-Perez, S., Villada, C., Pulopulos, M. M., Hidalgo, V., & Salvador, A. (2016). How are neuroticism and depression related to the psychophysiological stress response to acute stress in healthy older people? *Physiology & Behavior, 156,* 128-136.

*Quinn, M. E., Keil, D. C., Utke, S., & Joormann, J. (2014). Trait rumination moderates the effect of executive control training. *Journal of Experimental Psychopathology, 5*(3), 289-301.

*Richardson, C. M., Rice, K. G., & Devine, D. P. (2014). Perfectionism, emotion regulation, and the cortisol stress response. *Journal of Counseling Psychology*, *61*(1), 110-118.

Robinson, M. S., & Alloy, L. B. (2003). Negative cognitive styles and stress-reactive rumination interact to predict depression: A prospective study. *Cognitive Therapy and Research*, *27*(3), 275-291.

Ryan, R. (2014). Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Review Group reviews: Planning the analysis at protocol stage. *Cochrane Consumers and Communication Review Group.*

Sapolsky, R. M., Romero, L. M., & Munck, A. U. (2000). How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions. *Endocrine Reviews*, *21*(1), 55-89.

Schlotz, W., Hellhammer, J., Schulz, P., & Stone, A. A. (2004). Perceived work overload and chronic worrying predict weekend–weekday differences in the cortisol awakening response. *Psychosomatic Medicine*, *66*(2), 207-214.

Smith, S. M., & Vale, W. W. (2006). The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. *Dialogues in Clinical Neuroscience*, *8*(4), 383-395.

Smyth, J., Ockenfels, M. C., Porter, L., Kirschbaum, C., Hellhammer, D. H., & Stone, A. A. (1998). Stressors and mood measured on a momentary basis are associated with salivary cortisol secretion. *Psychoneuroendocrinology*, *23*(4), 353-370.

Taylor, S. E., Seeman, T. E., Eisenberger, N. I., Kozanian, T. A., Moore, A. N., & Moons, W. G. (2010). Effects of a supportive or an unsupportive audience on biological and psychological responses to stress. *Journal of Personality and Social Psychology, 98*(1), 47-56.

Tsigos, C., & Chrousos, G. P. (2002). Hypothalamic–pituitary–adrenal axis, neuroendocrine factors and stress. *Journal of Psychosomatic Research, 53*(4), 865-871.

Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., & Murray, C. J. (2004). Global burden of depressive disorders in the year 2000. *The British Journal of Psychiatry, 184*(5), 386-392.

*Vrshek-Schallhorn, S., Avery, B. M., Ditcheva, M., & Sapuram, V. R. (2018). The cortisol reactivity threshold model: Direction of trait rumination and cortisol reactivity association varies with stressor severity. *Psychoneuroendocrinology, 92*, 113-122.

*Vrshek-Schallhorn, S., Velkoff, E. A., & Zinbarg, R. E. (2018). Trait rumination and response to negative evaluative lab-induced stress: neuroendocrine, affective, and cognitive outcomes. *Cognition and Emotion*, 1-14.

*Way, B. M., & Taylor, S. E. (2010). [The serotonin transporter promoter polymorphism is associated with cortisol response to psychosocial stress]. Unpublished raw data.

Wetherell, M. A., Lovell, B., & Smith, M. A. (2015). The effects of an anticipated challenge on diurnal cortisol secretion. *Stress, 18*(1), 42-48.

Wiersma, J. E., van Oppen, P., Van Schaik, D. J., Van der Does, A. J., Beekman, A. T., & Penninx, B. W. (2011). Psychological characteristics of chronic depression: A longitudinal cohort study. *Journal of Clinical Psychiatry, 72*(3), 288-294.

*Wilson, M. C., Zilioli, S., Ponzi, D., Henry, A., Kubicki, K., Nickels, N., & Maestripieri, D. (2015). Cortisol reactivity to psychosocial stress mediates the relationship between extraversion and unrestricted sociosexuality. *Personality and Individual Differences, 86,* 427-431.

*Wirtz, P. H., Elsenbruch, S., Emini, L., Rüdisüli, K., Groessbauer, S., & Ehlert, U. (2007). Perfectionism and the cortisol response to psychosocial stress in men. *Psychosomatic Medicine, 69*(3), 249-255.

*Zoccola, P. M., & Dickerson, S. S. (2015). Extending the recovery window: Effects of trait rumination on subsequent evening cortisol following a laboratory performance stressor. *Psychoneuroendocrinology, 58,* 67-78.

*Zoccola, P. M., Dickerson, S. S., & Zaldivar, F. P. (2008). Rumination and cortisol responses to laboratory stressors. *Psychosomatic Medicine, 70*(6), 661-667.

*Zoccola, P. M., Quas, J. A., & Yim, I. S. (2010). Salivary cortisol responses to a psychosocial laboratory stressor and later verbal recall of the stressor: The role of trait and state rumination. *Stress, 13*(5), 435-443.

*Zoccola, P.M. (*n.d.*). Unpublished raw data.

*Zureck, E., Altstötter-Gleich, C., Wolf, O. T., & Brand, M. (2014). It depends: Perfectionism as a moderator of experimentally induced stress. *Personality and Individual Differences, 63,* 30-35.

APPENDIX A

TABLES AND FIGURES

Figure 1. Hypothesized Relationship Between Cortisol Reactivity and Depression Risk Factors Under Varying Levels of Stressor Severity

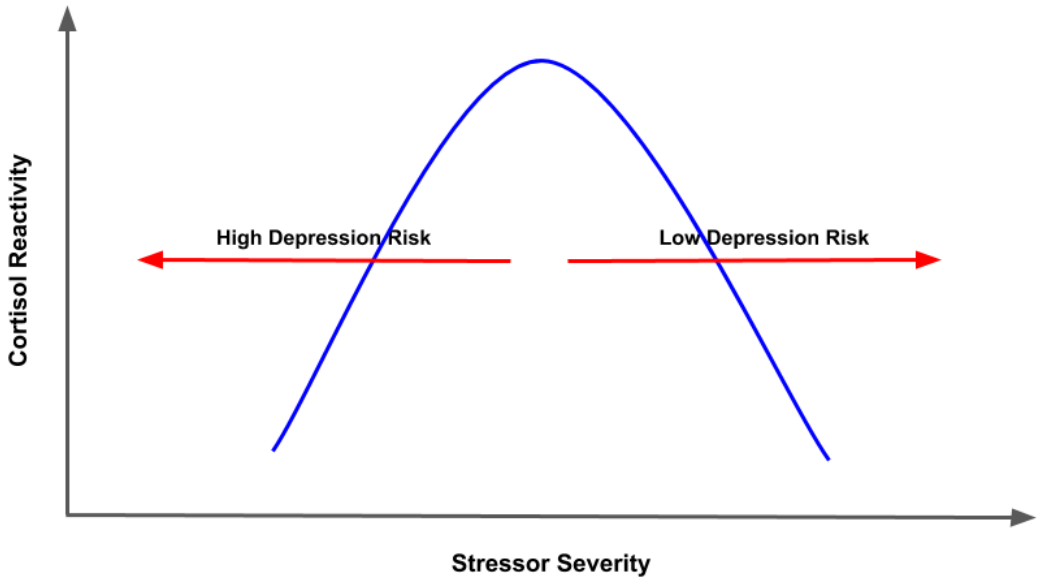
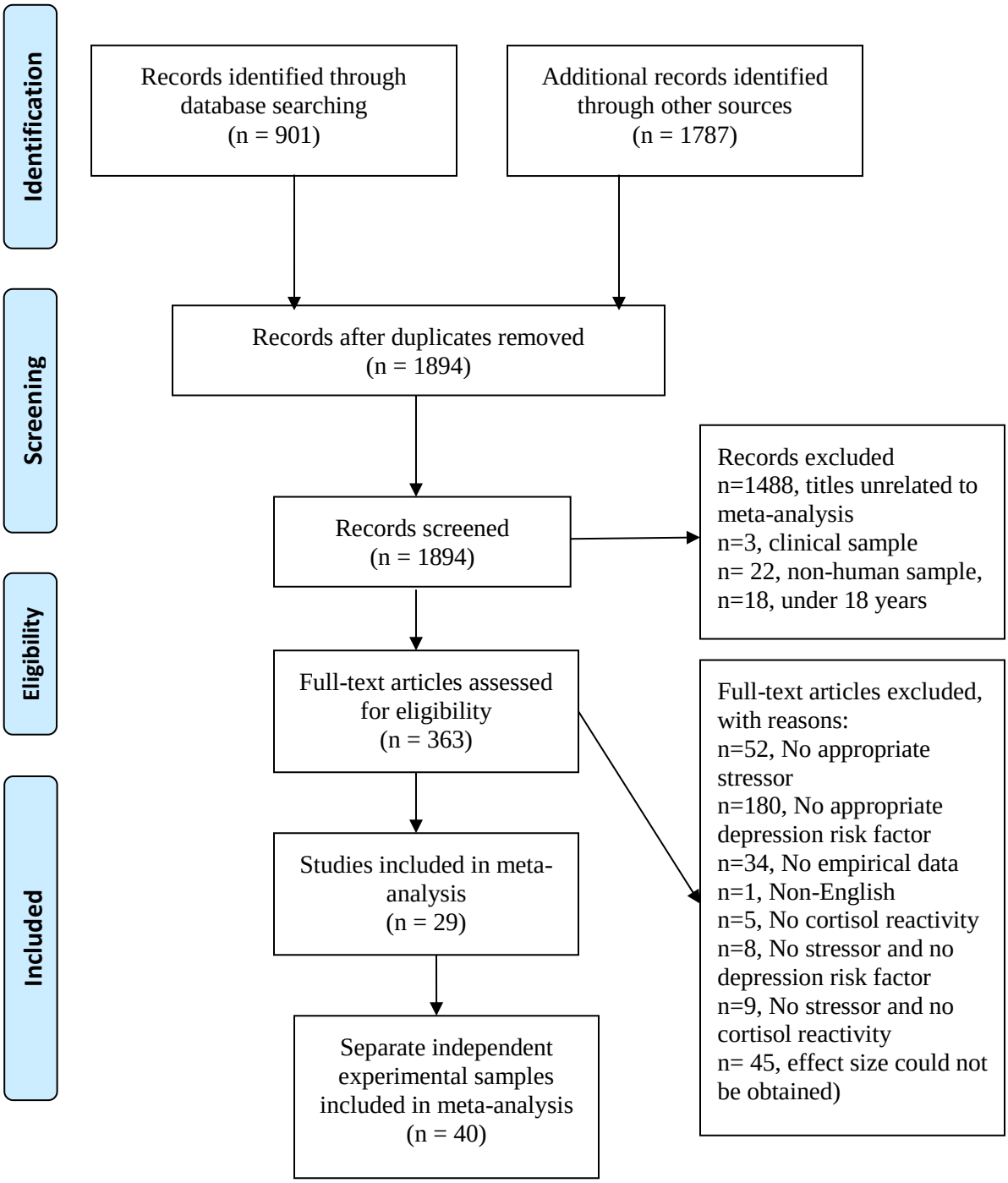


Table 1. Cortisol Reactivity Index Interpretation and Calculation

Cortisol Reactivity Index	**Interpretation**	**Derivation**
Area Under the Curve with respect to ground (AUCg) Pruessner, Kirschbaum, Meinlschmid, & Hellhammer (2003)	Total cortisol secretion over a period of time	$AUCg = \sum_{i=1}^{n-1} \frac{m_{(i+1)} + (m_i).\ t_i}{2}$ m = cortisol measurements t = time of measurement
Area under the curve with respect to increase (AUCi) Pruessner, Kirschbaum, Meinlschmid, & Hellhammer (2003)	Cortisol secretion over time that exceeds baseline levels, often used to operationalize reactivity	$AUCgi = AUCg - \left(m_1 . \sum_{i=1}^{n-1} . t_i\right)$ m = cortisol measurements t = time of measurement
Simple difference	Increase from a baseline measure (S1) to a later, sometimes peak, measure (S2)	S2 – S1
Quadratic effect	The extent of rise and fall of cortisol across time, as in a parabolic or inverted U-curve, often used to operationalize reactivity	T-values derived from growth curve models representing an interaction of the risk factor with time, e.g., Neuroticism x $Time^2$ predicting repeated measures of cortisol level

Figure 2. Flow Diagram of Included Studies and Reasons for Exclusion



Note. PRISMA Flow Diagram of Included Articles adapted from Moher et al., 2009

Table 2. Study Demographics

Author (Year)	**N**	**Mean Age (*SD*)**	**Proportion Female**	**Proportion Minority**
Bibbey (2013)*	352	58.23 (0.95)	0.54	0
Bibbey (2013)	352	58.23 (0.95)	0.54	0
Calingaert (2016)	42	22.95 (9.2)	0.67	0.67
Campos (2014)	59	20.68 (2.43)	1	0.8
Dinzeo (2015)	10	36.7 (8.77)	0.5	0.5
Dinzeo (2015)	10	36.7 (8.77)	0.5	0.5
Gerra (1998)	16	18.6 (0.5)	0.5	n/a
Gerra (2001)	20	25.6 (6.8)	0	0
Gianferante (2014)*	26	43.3 (19.18)	0.48	0.33
Gianferante (2014)*	26	43.3 (19.18)	0.48	0.33
Henckens (2016)*[C]	120	21.9 (2.62)	0	n/a
Henckens (2016)*[C]	120	21.9 (2.62)	0	n/a
Ioannou (2016)*[C]	16	49.4 (8)	0.44	n/a
Kirschbaum (1995)	20	22.4 (n/a)	0	n/a
Kirschbaum (1995)	20	22.4 (n/a)	0	n/a
Lewis (2018)*	33	37.73 (10.67)	0.39	0.59
Marin**[C]	16	22.1 (0.5)	0.56	n/a
Marin**[C]	16	22.1 (0.5)	0.56	n/a
McGirr & Turecki (2009)	16	44.18 (15.24)	0.63	0
Morris_A**[C]	53	22.97 (3.91)	0.62	0.15
Morris_B**[C]	49	22.97 (3.87)	0.65	0.25
Oswald (2006_A)	43	21.7 (2.8)	0	0.21

Oswald (2006_B)	25	21.4 (2.8)	1	0.4
Phillips (2005)	50	19.79 (2.25)	0.46	0.12
Phillips (2005)	48	19.79 (2.25)	0.46	0.12
Puig-Perez (2016_A)*	34	63.74 (0.705)	0.51	0
Puig-Perez (2016_B)*	36	64.66 (0.697)	0.56	0
Quinn (2014)	23	19.3 (1.49)	0.39	0.26
Richardson (2014)*	30	18.77 (0.824)	0.52	0.4
Richardson (2014)*	27	18.77 (0.824)	0.52	0.4
Richardson (2014)*	57	18.77 (0.824)	0.52	0.4
Richardson (2014)*	30	18.77 (0.824)	0.52	0.4
Richardson (2014)*	27	18.77 (0.824)	0.52	0.4
Richardson (2014)*	57	18.77 (0.824)	0.52	0.4
Richardson (2014)*	57	18.77 (0.824)	0.52	0.4
Richardson (2014)*	57	18.77 (0.824)	0.52	0.4
Vrshek-Schallhorn (2018_A1)	55	18.89 (1.048)	0.38	0.31
Vrshek-Schallhorn (2018_A2)	69	18.55 (0.718)	0.35	0.23
Vrshek-Schallhorn (2018_B1)*[C]	30	19.4 (1.59)	0.53	0.67
Vrshek-Schallhorn (2018_B1)*[C]	30	19.4 (1.59)	0.53	0.67
Vrshek-Schallhorn et al. (2018_B1)	30	19.4 (1.59)	0.53	0.67
Vrshek-Schallhorn (2018_B2)*[C]	24	20.23 (2.34)	0.5	0.73
Vrshek-Schallhorn (2018_B2)*[C]	24	20.23 (2.34)	0.5	0.73
Vrshek-Schallhorn (2018_B2)	26	20.23 (2.34)	0.5	0.73
Vrshek-Schallhorn (2018_B3)*[C]	29	19.28 (1.31)	0.45	0.62
Vrshek-Schallhorn (2018_B3)*[C]	29	19.28 (1.31)	0.45	0.62

Vrshek-Schallhorn (2018_B3)	29	19.28 (1.31)	0.45	0.62
Way & Taylor**[C]	59	18-35 (n/a)	0.61	0.73
Way & Taylor**[C]	59	18-35 (n/a)	0.61	0.73
Wilson (2015)	103	22.42 (3.9)	0	n/a
Wilson (2015)	103	22.42 (3.9)	0	n/a
Wirtz (2007)	42	42.5 (2)	0	0
Wirtz (2007)	42	42.5 (2)	0	0
Wirtz (2007)	42	42.5 (2)	0	0
Zoccola (*n.d.*)**	120	19.54 (2.08)	0.56	0.65
Zoccola (*n.d.*)**	120	19.54 (2.08)	0.56	0.65
Zoccola (*n.d.*)**	120	19.54 (2.08)	0.56	0.65
Zoccola (2008_A)*	28	21 (1.96)	0.64	0.57
Zoccola (2008_B)*	31	20.71 (1.44)	0.61	0.68
Zoccola (2010_A)*	59	20.1 (1.7)	0.53	0.55
Zoccola (2010_B)	13	19.77 (1.48)	0	0.54
Zoccola (2010_C)*	15	20.53 (1.81)	1	0.4
Zoccola (2015_A)*	87	19.82 (1.25)	0.49	0.75
Zoccola (2015_A)*	87	19.82 (1.25)	0.49	0.75
Zoccola (2015_B)*	57	19.86 (2.08)	0.51	0.79
Zoccola (2015_B)*	57	19.86 (2.08)	0.51	0.79
Zureck (2014_A)*[C]	42	23.94 (4.81)	0.75	n/a
Zureck (2014_A)*[C]	42	23.94 (4.81)	0.75	n/a
Zureck (2014_A)*[C]	42	23.94 (4.81)	0.75	n/a
Zureck (2014_B)*[C]	41	23.94 (4.81)	0.75	n/a

Zureck (2014_B)*[C]	41	23.94 (4.81)	0.75	n/a
Zureck (2014_B)*[C]	41	23.94 (4.81)	0.75	n/a

Note. In studies with multiple outcomes using the same participants, only the largest number of participants across the outcomes was used in calculating overall participant number (e.g. Richardson (2104), contributed 57 overall participants to this meta-analysis. Studies providing independent subgroups for analyses are designated with capital letter, A, B, C. * = unreported effect size of cortisol reactivity and depression risk factor, data obtained from author; ** = unpublished data; [C] = calculated effect size from raw provided data; n/a = not available

Table 3. Study Statistics and Characteristics

Author (Year)	N	Rept'ed r-stat	Rept'ed t-stat	Used r-stat	Used t-stat	Risk Factor	Social Eval	Hab	Un-cntrl	Aud Size	Antic	Cort Source	Time of day	React Index
Bibbey (2013)*	352	-0.095	-1.463	-0.095	-1.463	Extrav	Ambig	No	Yes	1	2	Saliva	PM	Simple Diff
Bibbey (2013)	352	-0.187	-3.2	-0.187	-3.2	Neurot	Ambig	No	Yes	1	2	Saliva	PM	Simple Diff
Calingaert (2016)	42		0.758		0.758	Neurot	Ambig	No	Yes	3	5	Saliva	PM	AUCg
Campos (2014)	59		2.38		2.38	Neurot	Ambig	No	Yes	2	3	Saliva	PM	Quadra tic
Dinzeo (2015)	10	-0.258		-0.258		Extrav	None	Yes	Yes	0	0	Saliva	PM	Simple Diff
Dinzeo (2015)	10	0.193		0.193		Neurot	None	Yes	Yes	0	0	Saliva	PM	Simple Diff
Gerra (1998)	16	0.078		0.078		Harm Av_Neurot	None	Yes	Yes	0	0	Plasma	PM	Simple Diff
Gerra (2001)	20	0.55		0.55		Harm Av_Neurot	Ambig	Yes	Yes	3	0	Plasma	PM	Simple Diff
Gianferante (2014)*	26	0.196		0.196		Rum1	Ambig	Yes	Yes	2	5	Saliva	PM	Simple Diff
Gianferante (2014)*	26	0.323		0.323		Rum2	Ambig	Yes	Yes	2	5	Saliva	PM	Simple Diff
Henckens (2016)*[C]	120	0.185		0.185		Extrav	None	Yes	Yes	0	0	Saliva	PM	AUCi
Henckens (2016)*[C]	120	0.108		0.108		Neurot	None	Yes	Yes	0	0	Saliva	PM	AUCi
Ioannou (2016)*[C]	16	0.077		0.077		Perf	Ambig	No	Yes	2	3	Saliva	PM	AUCi
Kirschbaum (1995)	20	0.67		0.67		Extrav	Ambig	Yes	Yes	3	10	Saliva	PM	AUCi
Kirschbaum (1995)	20	0.28		0.28		Neurot	Ambig	Yes	Yes	3	10	Saliva	PM	AUCi

Lewis (2018)*	33	0.09		0.09		Rum	Ambig	No	Yes	1	5	Saliva	PM	Simple Diff
Marin**[C]	16	0.065		0.065		Extrav	Ambig	No	Yes	2	10	Saliva	AM	AUCi
Marin**[C]	16	0.044		0.044		Neurot	Ambig	No	Yes	2	10	Saliva	AM	AUCi
McGirr & Turecki* (2009)	16	-0.025		-0.025		Perf	Ambig	No	Yes	1	5	Saliva	PM	AUCi
Morris_A**[C]	53	-0.319		-0.319		Neg Cog	Ambig	No	Yes	1	10	Saliva	PM	AUCi
Morris_B**[C]	49	-0.063		-0.063		Neg Cog	None	No	No	0	10	Saliva	PM	AUCi
Oswald (2006_A)	43	-0.31		-0.31		Extrav	Ambig	No	Yes	2	10	Plasma	PM	Other
Oswald (2006_B)	25	-0.389		-0.389		Neurot	Ambig	No	Yes	2	10	Plasma	PM	Other
Phillips (2005)	50	-0.31		-0.31		Neurot	Ambig	No	Yes	1	0	Saliva	PM	Simple Diff
Phillips (2005)	48	-0.29		-0.29		Neurot	Ambig	No	Yes	1	0	Saliva	PM	AUCi
Puig-Perez (2016_A)*	34	0.163		0.163		Neurot	None	No	No	0	3	Saliva	PM	AUCg
Puig-Perez (2016_B)*	36	0.044		0.044		Neurot	Ambig	No	Yes	2	3	Saliva	PM	AUCg
Quinn (2014)	23		4.32		4.32	Rum	Ambig	No	Yes	1	3	Saliva	AM/PM	Simple Diff
Richardson (2014)*	30	0.066		0.066		Extrav 1	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	27	0.01		0.01		Extrav 2	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	57	0.039		0.039		Extrav 3	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	30	0.148		0.148		Neurot 1	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	27	-0.264		-0.264		Neurot 2	Ambig	No	Yes	3	6	Saliva	PM	AUCi

Richardson (2014)*	57	-0.011		-0.011		Neurot 3	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	57	-0.124		-0.124		Perf1	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Richardson (2014)*	57	-0.198		-0.198		Perf2	Ambig	No	Yes	3	6	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_A1)	55		-1.72		-1.72	Rum	Neg	No	Yes	2	5	Saliva	PM	Quadratic
Vrshek-Schallhorn (2018_A2)	69		1.2		1.2	Rum	None	No	Yes	0	5	Saliva	PM	Quadratic
Vrshek-Schallhorn (2018_B1)*[C]	30	0.01		0.01		Extrav	Neg	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B1)*[C]	30	-0.53		-0.53		Neurot	Neg	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn et al. (2018_B1)	30	-0.379		-0.379		Rum	Neg	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B2)*[C]	24	0.158		0.158		Extrav	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B2)*[C]	24	-0.025		-0.025		Neurot	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B2)	26	0.4		0.4		Rum	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B3)*[C]	29	-0.045		-0.045		Extrav	None	No	No	0	5	Saliva	PM	AUCi
Vrshek-Schallhorn (2018_B3)*[C]	29	-0.29		-0.29		Neurot	None	No	No	0	5	Saliva	PM	AUCi

Vrshek-Schallhorn (2018_B3)	29	-0.111		-0.111		Rum	None	No	No	0	5	Saliva	PM	AUCi
Way & Taylor**[C]	59	0.076		0.076		Extrav	Neg	No	Yes	2	5	Saliva	PM	AUCi
Way & Taylor**[C]	59	-0.27		-0.27		Neurot	Neg	No	Yes	2	5	Saliva	PM	AUCi
Wilson (2015)	103	0.199		0.199		Extrav 1	Ambig	No	Yes	2	3	Saliva	PM	Simple Diff
Wilson (2015)	103	0.212		0.212		Extrav 2	Ambig	No	Yes	2	3	Saliva	PM	Simple Diff
Wirtz (2007)	42	0.321		0.321		Extrav	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Wirtz (2007)	42	0.289		0.289		Neurot	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Wirtz (2007)	42	0.322		0.322		Perf	Ambig	No	Yes	2	5	Saliva	PM	AUCi
Zoccola (*n.d.*)**	120		-0.87		-0.87	Neurot	Ambig	No	Yes	2	10	Saliva	PM	Quadratic
Zoccola (*n.d.*)**	120		-3.75		-3.75	Rum1	Ambig	No	Yes	2	10	Saliva	PM	Quadratic
Zoccola (*n.d.*)**	120		-2.37		-2.37	Rum2	Ambig	No	Yes	2	10	Saliva	PM	Quadratic
Zoccola (2008_A)*	28		-1.95		-1.95	Rum	Ambig	No	Yes	2	10	Saliva	PM	Quadratic
Zoccola (2008_B)*	31		-0.62		-0.62	Rum	None	No	No	0	10	Saliva	PM	Quadratic
Zoccola (2010_A)*	59		2.6		2.6	Rum	Ambig	No	Yes	2	3	Saliva	PM	Quadratic
Zoccola (2010_B)	13		2.62		2.62	Rum	Ambig	Yes	Yes	1	0	Saliva	PM	Quadratic
Zoccola (2010_C)*	15		0.6		0.6	Rum	Ambig	Yes	Yes	1	0	Saliva	PM	Quadratic
Zoccola (2015_A)*	87		-0.81		-0.81	Neurot	Ambig	No	Yes	2	5	Saliva	PM	Quadratic
Zoccola (2015_A)*	87		-0.53		-0.53	Rum	Ambig	No	Yes	2	5	Saliva	PM	Quadratic

Zoccola (2015_B)*	57		-0.12		-0.12	Neurot	None	No	Yes	0	5	Saliva	PM	Quadratic
Zoccola (2015_B)*	57		-0.004		-0.004	Rum	None	No	Yes	0	5	Saliva	PM	Quadratic
Zureck (2014_A)*[C]	42	-0.27		-0.27		Neurot	Ambig	No	Yes	2	5	Saliva	AM/PM	AUCi
Zureck (2014_A)*[C]	42	0.008		0.008		Perf1	Ambig	No	Yes	2	5	Saliva	AM/PM	AUCi
Zureck (2014_A)*[C]	42	-0.061		-0.061		Perf2	Ambig	No	Yes	2	5	Saliva	AM/PM	AUCi
Zureck (2014_B)*[C]	41	0.087		0.087		Neurot	None	No	No	0	5	Saliva	AM/PM	AUCi
Zureck (2014_B)*[C]	41	0.034		0.034		Perf1	None	No	No	0	5	Saliva	AM/PM	AUCi
Zureck (2014_B)*[C]	41	0.161		0.161		Perf2	None	No	No	0	5	Saliva	AM/PM	AUCi

Note. Studies with no year represent unpublished data. Studies providing independent subgroups for analyses are designated with capital letter, A, B, C; * = unreported effect size for cortisol reactivity and depression risk factor, data obtained from author; ** = unpublished data; [C] = calculated effect size from raw provided data; rept'ed = reported; n/a = not available; extrav = extraversion, neg cog = negative cognitive style, neurot = neuroticism; harm av = harm avoidance; perf = perfectionism; rum = rumination; ambig = ambiguous; none = no social evaluation; neg = negative social evaluation; social eval = social evaluation; uncntrl = uncontrollability; hab = habituation; antic = anticipation; aud size = audience size; cort source = cortisol source; react index = reactivity index; simple diff = simple difference.

Table 4. Overall Effect Size By Depression Risk Factor

		Effect Size			Heterogeneity	
	N	**Hedges's *g***	**95% *CI***	***p*-value**	***I*²**	***p*-value**
Combined Depression Risk Factors	40	0.039	-0.112, 0.190	0.609	55.813	0.000**
Extraversion	13	0.152	-0.089, 0.393	0.216	54.215	0.010**
Negative Cognitive Style	2	-0.392	-0.92, 0.136	0.146	41.148	0.192
Neuroticism	24	-0.059	-0.24, 0.123	0.527	53.588	0.001**
Perfectionism	6	0.075	-0.225, 0.376	0.623	12.108	0.338
Rumination	16	0.077	-0.214, 0.368	0.604	67.096	0.000**

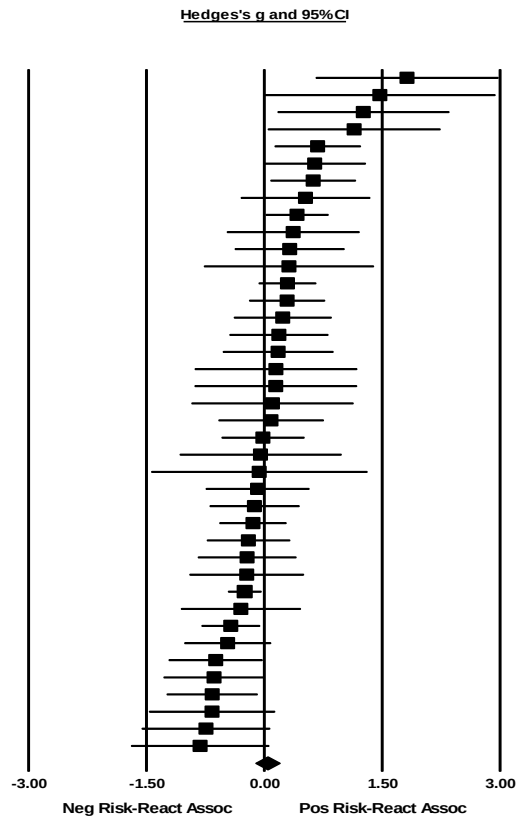
Note. *=significant at p=/<0.05 level; **=significant at p=/<0.01 Individual depression risk factors do not add up to the total overall studies because in studies with multiple outcomes, effect sizes are pooled.

Figure 3. Forest Plot of Included Studies

a)

Depression Risk-Reactivity Effect Size

Study name	Hedges's g	Outcome	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Quinn (2014)	1.817	Rumination	0.592	0.351	0.656	2.978	3.068	0.002
Zoccola (2010_B)	1.470	Rumination	0.750	0.562	0.000	2.939	1.960	0.050
Gerra et al. (2001)	1.261	Harm Av_Neurot	0.556	0.309	0.171	2.352	2.268	0.023
Kirschbaum (1995)	1.144	Combined	0.559	0.313	0.047	2.240	2.045	0.041
Zoccola (2010_A)	0.680	Rumination	0.279	0.078	0.133	1.226	2.437	0.015
Wirtz (2007)	0.642	Combined	0.331	0.109	-0.006	1.290	1.941	0.052
Campos (2014)	0.622	Neuroticism	0.277	0.076	0.080	1.164	2.250	0.024
Gianferante (2014)	0.524	Combined	0.419	0.176	-0.298	1.346	1.250	0.211
Wilson (2015)	0.417	Combined	0.203	0.041	0.019	0.814	2.055	0.040
Vrshek-Schallhorn (2018_B2)	0.369	Combined	0.430	0.185	-0.474	1.211	0.858	0.391
Puig-Perez (2016_A)	0.323	Neuroticism	0.355	0.126	-0.374	1.019	0.908	0.364
Zoccola (2010_C)	0.313	Rumination	0.551	0.303	-0.766	1.393	0.569	0.570
Henckens (2016)	0.295	Combined	0.186	0.035	-0.069	0.659	1.587	0.113
Vrshek-Schallhorn (2018_A2)	0.290	Rumination	0.246	0.061	-0.192	0.772	1.178	0.239
Calingaert (2016)	0.235	Neuroticism	0.316	0.100	-0.385	0.855	0.743	0.457
Zureck (2014_B)	0.186	Combined	0.320	0.102	-0.441	0.813	0.581	0.561
Lewis (2018)	0.176	Rumination	0.358	0.128	-0.525	0.877	0.493	0.622
Gerra et al. (1998)	0.148	Harm Av_Neurot	0.526	0.277	-0.883	1.179	0.281	0.779
Ioannou (2016)	0.146	Perfectionism	0.526	0.277	-0.885	1.177	0.278	0.781
Marin	0.103	Combined	0.525	0.276	-0.926	1.133	0.197	0.844
Puig-Perez (2016_B)	0.086	Neuroticism	0.341	0.116	-0.582	0.754	0.253	0.800
Zoccola (2015_B)	-0.016	Combined	0.268	0.072	-0.543	0.510	-0.061	0.951
McGirr (2009)	-0.047	Perfectionism	0.525	0.275	-1.075	0.981	-0.090	0.928
Dinzeo (2015)	-0.064	Combined	0.701	0.492	-1.438	1.311	-0.091	0.928
Richardson (2014)	-0.085	Combined	0.336	0.113	-0.743	0.573	-0.254	0.800
Morris_B	-0.124	Negative Cognitive Style	0.291	0.085	-0.694	0.446	-0.427	0.669
Zoccola (2015_A)	-0.144	Combined	0.217	0.047	-0.569	0.281	-0.664	0.507
Way & Taylor	-0.201	Combined	0.269	0.072	-0.729	0.326	-0.748	0.454
Zureck (2014_A)	-0.218	Combined	0.319	0.101	-0.842	0.406	-0.685	0.493
Zoccola (2008_B)	-0.224	Rumination	0.371	0.137	-0.950	0.502	-0.605	0.545
Bibbey (2013)	-0.249	Combined	0.108	0.012	-0.460	-0.037	-2.308	0.021
Vrshek-Schallhorn (2018_B3)	-0.298	Combined	0.388	0.150	-1.058	0.462	-0.768	0.443
Zoccola	-0.426	Combined	0.189	0.036	-0.797	-0.056	-2.256	0.024
Vrshek-Schallhorn (2018_A1)	-0.466	Rumination	0.281	0.079	-1.016	0.085	-1.658	0.097
Phillips (2005)	-0.619	Combined	0.304	0.093	-1.215	-0.023	-2.035	0.042
Oswald (2006_A)	-0.640	Extraversion	0.326	0.107	-1.280	-0.000	-1.961	0.050
Morris_A	-0.663	Negative Cognitive Style	0.294	0.086	-1.239	-0.087	-2.256	0.024
Vrshek-Schallhorn (2018_B1)	-0.665	Combined	0.408	0.166	-1.464	0.135	-1.629	0.103
Zoccola (2008_A)	-0.743	Rumination	0.416	0.173	-1.557	0.072	-1.786	0.074
Oswald (2006_B)	-0.817	Neuroticism	0.448	0.200	-1.694	0.061	-1.825	0.068
	0.039		0.077	0.006	-0.112	0.190	0.512	0.609



Note. Squares represent effect size of each experimental sample, diamond represents summary effect size.

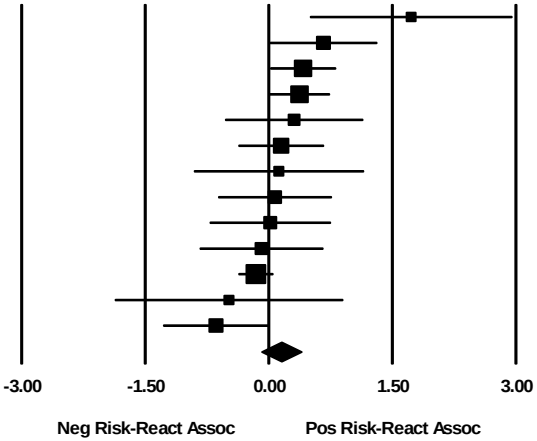
b)

Extraversion-Reactivity Effect Size

Study name	Hedges's g	Outcome	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Kirschbaum (1995)	1.729	Extraversion	0.626	0.392	0.502	2.955	2.762	0.006
Wirtz (2007)	0.665	Extraversion	0.332	0.110	0.015	1.315	2.005	0.045
Wilson (2015)	0.417	Combined	0.203	0.041	0.019	0.814	2.055	0.040
Henckens (2016)	0.374	Extraversion	0.187	0.035	0.008	0.741	2.001	0.045
Vrshek-Schallhorn (2018_B2)	0.309	Extraversion	0.427	0.182	-0.527	1.145	0.724	0.469
Way & Taylor	0.150	Extraversion	0.264	0.070	-0.368	0.669	0.569	0.570
Marin	0.123	Extraversion	0.526	0.276	-0.907	1.153	0.234	0.815
Richardson (2014)	0.075	Combined	0.351	0.123	-0.613	0.763	0.214	0.831
Vrshek-Schallhorn (2018_B1)	0.019	Extraversion	0.375	0.140	-0.715	0.753	0.052	0.959
Vrshek-Schallhorn (2018_B3)	-0.088	Extraversion	0.382	0.146	-0.836	0.660	-0.229	0.819
Bibbey (2013)	-0.156	Extraversion	0.107	0.011	-0.366	0.054	-1.456	0.145
Dinzeo (2015)	-0.482	Extraversion	0.707	0.499	-1.868	0.903	-0.683	0.495
Oswald (2006_A)	-0.640	Extraversion	0.326	0.107	-1.280	-0.000	-1.961	0.050
	0.152		0.123	0.015	-0.089	0.393	1.238	0.216

Statistics for each study

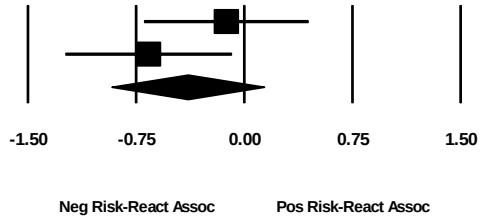
Hedges's g and 95% CI



c)

Neg Cog Style-Reactivity Effect Size

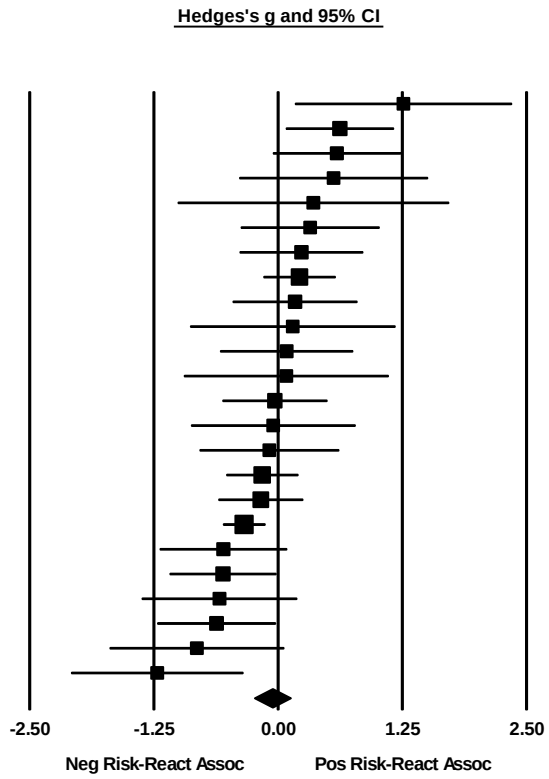
Study name	Outcome	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Morris_B	Negative Cognitive Style	-0.124	0.291	0.085	-0.694	0.446	-0.427	0.669
Morris_A	Negative Cognitive Style	-0.663	0.294	0.086	-1.239	-0.087	-2.256	0.024
		-0.392	0.269	0.073	-0.920	0.136	-1.454	0.146



d)

Neuroticism-Reactivity Effect Size

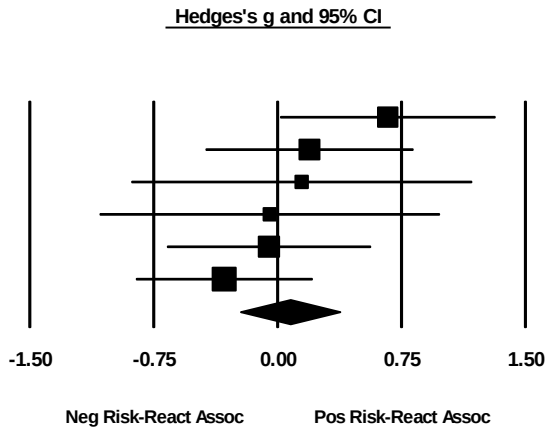
Study name	Hedges's g	Outcome	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Gerra et al. (2001)	1.261	Harm Av_Neurot	0.556	0.309	0.171	2.352	2.268	0.023
Campos (2014)	0.622	Neuroticism	0.277	0.076	0.080	1.164	2.250	0.024
Wirtz (2007)	0.592	Neuroticism	0.328	0.108	-0.051	1.236	1.805	0.071
Kirschbaum (1995)	0.559	Neuroticism	0.484	0.234	-0.390	1.507	1.154	0.248
Dinzeo (2015)	0.355	Neuroticism	0.696	0.484	-1.009	1.719	0.511	0.610
Puig-Perez (2016_A)	0.323	Neuroticism	0.355	0.126	-0.374	1.019	0.908	0.364
Calingaert (2016)	0.235	Neuroticism	0.316	0.100	-0.385	0.855	0.743	0.457
Henckens (2016)	0.216	Neuroticism	0.185	0.034	-0.146	0.578	1.168	0.243
Zureck (2014_B)	0.171	Neuroticism	0.319	0.102	-0.455	0.797	0.536	0.592
Gerra et al. (1998)	0.148	Harm Av_Neurot	0.526	0.277	-0.883	1.179	0.281	0.779
Puig-Perez (2016_B)	0.086	Neuroticism	0.341	0.116	-0.582	0.754	0.253	0.800
Marin	0.083	Neuroticism	0.525	0.276	-0.946	1.112	0.159	0.874
Zoccola (2015_B)	-0.032	Neuroticism	0.268	0.072	-0.558	0.494	-0.119	0.905
Vrshek-Schallhorn (2018_B2)	-0.048	Neuroticism	0.422	0.178	-0.874	0.778	-0.115	0.909
Richardson (2014)	-0.087	Combined	0.358	0.128	-0.788	0.614	-0.244	0.808
Zoccola	-0.159	Neuroticism	0.184	0.034	-0.520	0.202	-0.864	0.388
Zoccola (2015_A)	-0.174	Neuroticism	0.217	0.047	-0.600	0.251	-0.802	0.422
Bibbey (2013)	-0.341	Neuroticism	0.108	0.012	-0.554	-0.129	-3.150	0.002
Zureck (2014_A)	-0.550	Neuroticism	0.326	0.106	-1.190	0.089	-1.686	0.092
Way & Taylor	-0.553	Neuroticism	0.274	0.075	-1.090	-0.017	-2.020	0.043
Vrshek-Schallhorn (2018_B3)	-0.589	Neuroticism	0.398	0.159	-1.370	0.192	-1.479	0.139
Phillips (2005)	-0.619	Combined	0.304	0.093	-1.215	-0.023	-2.035	0.042
Oswald (2006_B)	-0.817	Neuroticism	0.448	0.200	-1.694	0.061	-1.825	0.068
Vrshek-Schallhorn (2018_B1)	-1.216	Neuroticism	0.442	0.195	-2.082	-0.351	-2.754	0.006
	-0.059		0.093	0.009	-0.240	0.123	-0.633	0.527



e)

Perfectionism-Reactivity Effect Size

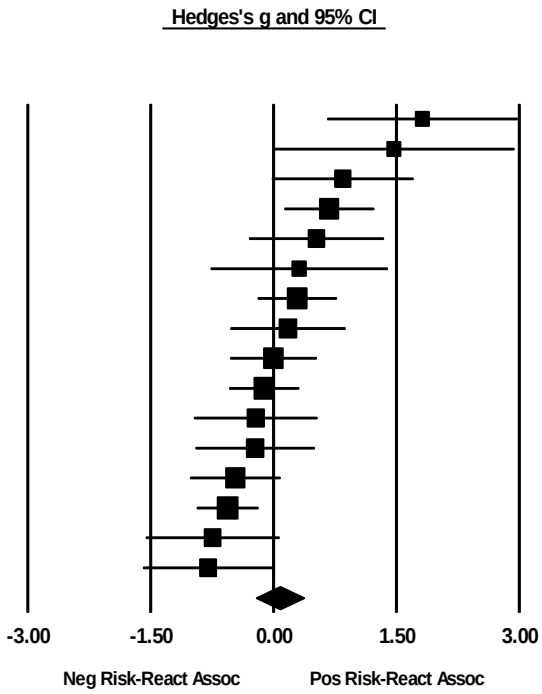
Study name	Outcome	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Wirtz (2007)	Perfectionism	0.667	0.332	0.110	0.017	1.318	2.011	0.044
Zureck (2014_B)	Combined	0.193	0.320	0.103	-0.435	0.821	0.603	0.546
Ioannou (2016)	Perfectionism	0.146	0.526	0.277	-0.885	1.177	0.278	0.781
McGirr (2009)	Perfectionism	-0.047	0.525	0.275	-1.075	0.981	-0.090	0.928
Zureck (2014_A)	Combined	-0.052	0.315	0.099	-0.669	0.564	-0.166	0.868
Richardson (2014)	Combined	-0.322	0.272	0.074	-0.856	0.211	-1.185	0.236
		0.075	0.153	0.023	-0.225	0.376	0.492	0.623



f)

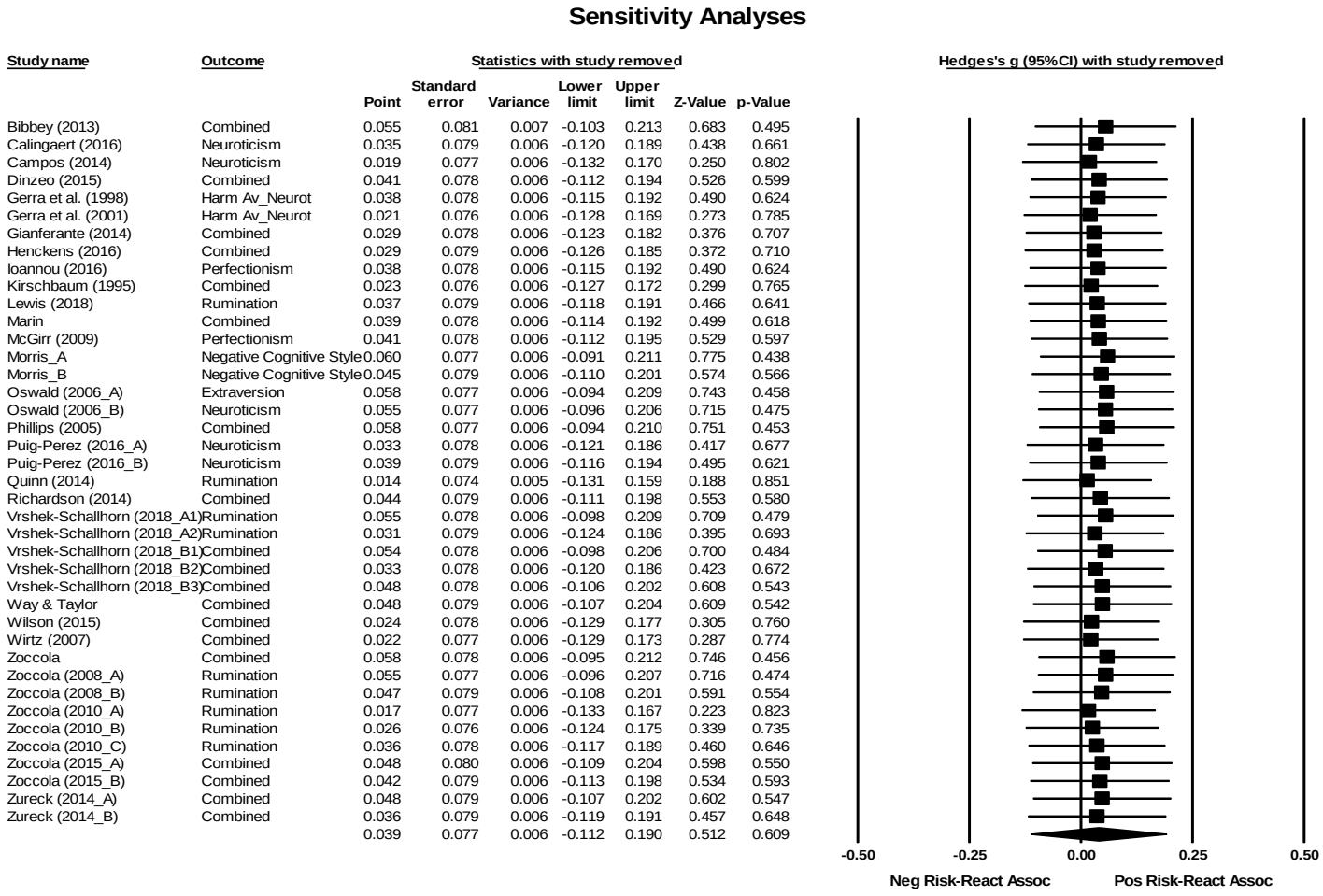
Rumination-Reactivity Effect Size

Study name	Hedges's g	Outcome	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Quinn (2014)	1.817	Rumination	0.592	0.351	0.656	2.978	3.068	0.002
Zoccola (2010_B)	1.470	Rumination	0.750	0.562	0.000	2.939	1.960	0.050
Vrshek-Schallhorn (2018_B2)	0.845	Rumination	0.441	0.194	-0.018	1.709	1.918	0.055
Zoccola (2010_A)	0.680	Rumination	0.279	0.078	0.133	1.226	2.437	0.015
Gianferante (2014)	0.524	Combined	0.419	0.176	-0.298	1.346	1.250	0.211
Zoccola (2010_C)	0.313	Rumination	0.551	0.303	-0.766	1.393	0.569	0.570
Vrshek-Schallhorn (2018_A2)	0.290	Rumination	0.246	0.061	-0.192	0.772	1.178	0.239
Lewis (2018)	0.176	Rumination	0.358	0.128	-0.525	0.877	0.493	0.622
Zoccola (2015_B)	-0.001	Rumination	0.268	0.072	-0.527	0.525	-0.004	0.997
Zoccola (2015_A)	-0.114	Rumination	0.217	0.047	-0.539	0.311	-0.526	0.599
Vrshek-Schallhorn (2018_B3)	-0.217	Rumination	0.384	0.147	-0.969	0.535	-0.566	0.571
Zoccola (2008_B)	-0.224	Rumination	0.371	0.137	-0.950	0.502	-0.605	0.545
Vrshek-Schallhorn (2018_A1)	-0.466	Rumination	0.281	0.079	-1.016	0.085	-1.658	0.097
Zoccola	-0.560	Combined	0.191	0.037	-0.935	-0.185	-2.927	0.003
Zoccola (2008_A)	-0.743	Rumination	0.416	0.173	-1.557	0.072	-1.786	0.074
Vrshek-Schallhorn (2018_B1)	-0.797	Rumination	0.405	0.164	-1.590	-0.004	-1.969	0.049
	0.077		0.148	0.022	-0.214	0.368	0.519	0.604



Note. Effect sizes are ordered from largest to smallest in figure 3a-f.

Figure 4. Sensitivity Analyses



Note. Effect sizes are ordered from largest to smallest.

Table 5. Continuous Moderators

Predictor	N	Coefficient	95% CI	P-value
Combined Depression Risk Factor	40			
Stressor Severity Lnr		0.254	-0.379, 0.888	0.431
Stressor Severity Qdrtc		-0.251	-0.602, 0.101	0.163
Extraversion	13			
Stressor Severity Lnr		0.173	-0.976, 1.322	0.768
Stressor Severity Qdrtc		-0.090	-0.644, 0.465	0.751
Neuroticism	24			
Stressor Severity Lnr		0.207	-0.505, 0.919	0.570
Stressor Severity Qdrtc		-0.321	-0.723, 0.081	0.118
Rumination	16			
Stressor Severity Lnr		0.888	-0.318, 2.094	0.149
Stressor Severity Qdrtc		-0.593	-1.216, 0.031	0.062
Mean Age	39	0.003	-0.009, 0.015	0.620
Proportion Female	40	-0.598	-1.164, -0.033	0.038*
Proportion Minority	32	-0.074	-0.713, 0.565	0.821
Audience Size	40	0.030	-0.13, 0.19	0.713
Anticipation	40	-0.069	-0.114, -0.023	0.003**
Task Time	40	0.021	-0.008, 0.05	0.154
TaskAnticip	40	-0.003	-0.036, 0.029	0.844

*Note.**=significant at p=/<0.05 level; **=significant at p=/<0.01 level; Lnr=linear, Qdrtc=quadratic; TaskAnticip=combined anticipation time and task time

Table 6. Categorical Moderators

	N	Hedge's *g*	95% Confidence Interval	*p*-value	*Q*-value	*p*-value
Cortisol Source						
Plasma	4	-0.169	-0.722, 0.384	0.549		
Saliva	36	0.057	-0.101, 0.214	0.481		
Total between					0.591	0.442
Overall	40	0.040	-0.112, 0.191	0.608		
Depression Risk						
Extraversion	16	0.169	-0.056, 0.393	0.140		
Negative Cognitive Style	2	-0.392	-0.995, 0.211	0.202		
Neuroticism	27	-0.080	-0.256, 0.095	0.370		
Perfectionism	9	0.027	-0.28, 0.334	0.862		
Rumination	18	0.013	-0.206, 0.232	0.904		
Total between					4.646	0.326
Overall	72	0.001	-0.159, 0.161	0.992		
Habituation						
Habituation	8	0.541	0.151, 0.931	0.007*		
No Habituation	32	-0.043	-0.196, 0.110	0.583		
Total between					7.455	0.006**
Overall	40	0.220	-0.349, 0.79	0.448		
Reactivity Index						
AUCg	3	0.214	-0.336, 0.764	0.446		
AUCi	15	0.005	-0.251, 0.26	0.972		
Other	2	-0.713	-1.431, 0.004	0.051		
Quadratic	11	0.034	-0.246, 0.315	0.811		
Simple Diff	8	0.345	-0.016, 0.707	0.061		
Total between					7.376	0.117
Overall	39	0.043	-0.233, 0.319	0.760		
Statistic Type						
r	25	0.045	-0.152, 0.243	0.652		
rp	1	0.642	-0.295, 1.578	0.179		
t	11	0.031	-0.243, 0.304	0.826		
Total between					1.545	0.462
Overall	37	0.066	-0.132, 0.263	0.515		
Time of Day						
AM	1	0.103	-1.129, 1.335	0.870		
AM/PM	3	0.333	-0.254, 0.919	0.266		
PM	36	0.018	-0.142, 0.177	0.828		
Total between					1.042	0.594
Overall	40	0.045	-0.124, 0.214	0.602		
Uncontrollability						
Controllable	5	-0.021	-0.448, 0.406	0.924		
Uncontrollable	35	0.049	-0.114, 0.213	0.555		
Total between					0.090	0.764
Overall	40	0.040	-0.113, 0.193	0.605		

Note. *=Significant at p=/<0.05 level, **=significant at p=/<0.01

Figure 5. Publication Bias Funnel Plot

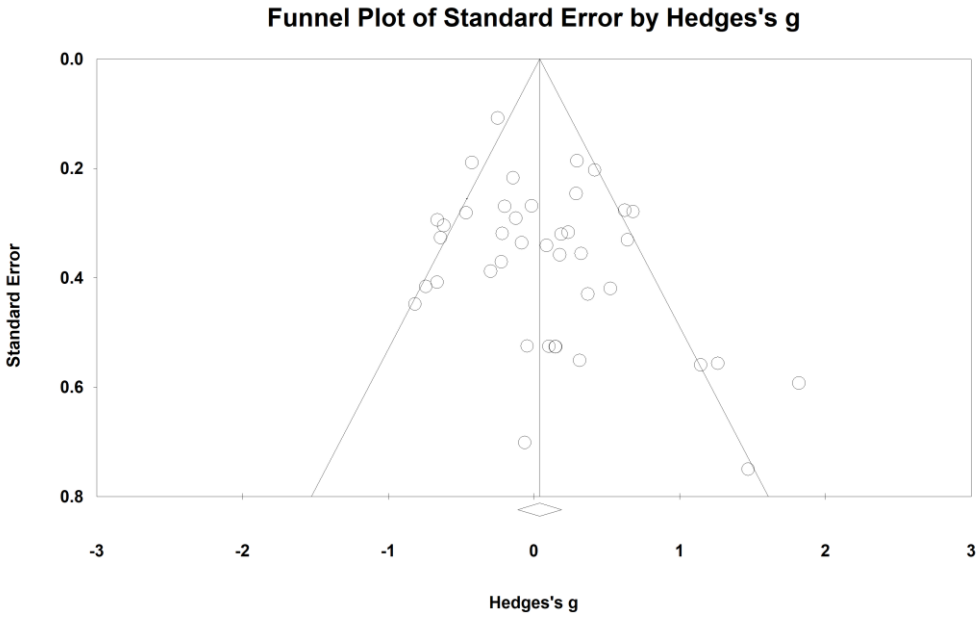


Table 7. Example of Behavioral Descriptors for Varying Levels of Severity

Intended Social Evaluation	**Behavioral Descriptors**
No Social Evaluation	- No evaluative audience present - No direct eye-contact if experimenter present - Told performance will not be evaluated - Neutral in demeanor, no explicit negative or positive verbal or nonverbal behaviors.
Ambiguous Social Evaluation	- Presence of evaluative audience - Direct eye contact - Told performance will be evaluated - Neutral audience tone, intended to convey no overt positive or negative feedback (neutral or strict tone, no smiling or nodding, no criticism or support of performance)
Negative Social Evaluation	- Presence of evaluative audience - Direct eye contact - Told performance will be evaluated - Provision of explicit negative evaluation such as criticism or harassment of performance - Body language communicating dissatisfaction with performance (e.g., shaking head, appearing bored, not paying attention)
Positive Social Evaluation	- Presence of evaluative audience - Direct eye contact - Told performance will be evaluated - Provision of explicit positive evaluation (e.g., nodding, smiling)